



Contribution de l'albumen au développement de la graine chez *Medicago truncatula*: caractérisation d'un facteur de transcription de type DOF exprimé dans l'albumen chalazal

Mélanie Noguero

► To cite this version:

Mélanie Noguero. Contribution de l'albumen au développement de la graine chez *Medicago truncatula*: caractérisation d'un facteur de transcription de type DOF exprimé dans l'albumen chalazal. Sciences agricoles. Université de Bourgogne, 2014. Français. NNT : 2014DIJOS022 . tel-01155045

HAL Id: tel-01155045

<https://theses.hal.science/tel-01155045>

Submitted on 26 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITE DE BOURGOGNE

Ecole Doctorale Environnements Santé n°554



Institut National de la Recherche Agronomique

UMR 1347 Agroécologie - pôle GEAPSI



Agroécologie
Dijon
Unité de Recherche



THÈSE

Pour l'obtention du grade de

Docteur de l'Université de Bourgogne

Discipline : Sciences Vie

Spécialité: Biologie moléculaire végétale

Contribution de l'albumen au développement de la graine
chez *Medicago truncatula* : caractérisation d'un facteur de
transcription de type DOF exprimé dans l'albumen chalazal.

par Mélanie NOGUERO

soutenue le 23 Mai 2014 devant la commission d'examen:

Gwyneth INGRAM, Chargée de recherches, CNRS, ENS Lyon

Rapporteur

Loic RAJJOU, Maître de conférences, AgroParisTech

Rapporteur

Nahalie LEBORGNE-CASTEL, Professeur Universités, Dijon

Examineur

Bertrand DUBREUCQ, Directeur de recherches, INRA, AgroParisTech

Examineur

Karine GALLARDO, Chargée de recherches, INRA, Dijon

Co-encadrant de thèse

Richard THOMPSON, Directeur de recherches, INRA, Dijon

Directeur de thèse

Résumé

Dans un contexte actuel qui tend vers une réduction d'intrants dans les systèmes de culture et une relance de la production de protéines végétales pour réduire la dépendance alimentaire de la France, la culture des légumineuses constitue une alternative. Les légumineuses à graines offrent une source riche en protéines pour l'alimentation animale et humaine. Au sein de l'UMR1347 Agroécologie, le pôle "déterminismes Génétiques et Environnementaux de l'Adaptation des Plantes à des Systèmes de culture Innovants" (GEAPSI) étudie *via* une approche multidisciplinaire (génétique, écophysiologie, physiologie moléculaire) l'adaptation des espèces végétales, et notamment des légumineuses aux contraintes environnementales. Ces travaux de thèse ont été réalisés au sein de l'équipe "Étude des Mécanismes Moléculaires" qui s'intéresse plus particulièrement à des critères de qualité de la graine (teneur en protéine, taille de graine) et au déterminisme génétique de ces caractères chez la plante modèle *Medicago truncatula* en vue d'un transfert de connaissances vers l'espèce cible *Pisum sativum*.

Chez les légumineuses, la taille de la graine est déterminée par la capacité de l'embryon à se diviser lors de l'embryogenèse et à accumuler des réserves lors du remplissage. Aux stades précoces du développement, l'assimilation de nutriments est réalisée majoritairement par les tissus qui entourent l'embryon: l'albumen et le tégument. Les recherches menées visent à évaluer la contribution de l'albumen dans le développement de la graine chez *M. truncatula*. Nous avons révélé plusieurs gènes DOF (DNA-binding with One zinc Finger) comme étant exprimé dans ce tissu. Ils appartiennent à une large famille de facteurs de transcription impliqués dans de nombreux processus biologiques mais dont les rôles restent à préciser. Un de ces gènes, nommé *DASH* pour *Dof Affecting Seed embryogenesis and Hormone metabolism*, est exprimé préférentiellement dans l'albumen lors de l'embryogenèse. Des mutants TILLING et Tnt-1 isolés pour ce facteur de transcription sont affectés dans le développement de la graine (avortement à ≈ 10 jap). La cytologie du développement aux stades précoces (6 à 10 jap) a révélé que l'expression de ce gène dans l'albumen est requise pour un développement normal de l'embryon, démontrant le rôle de l'albumen dans le contrôle de l'embryogenèse chez les légumineuses. Une étude comparative du transcriptome des gousses de *dash* vs sauvage a permis d'émettre des hypothèses quant à la fonction du gène *DASH*. Une dérégulation du métabolisme hormonal, en particulier de l'auxine, a été mise en évidence et plusieurs gènes cibles potentiels de ce facteur de transcription ont été sélectionnés. Une comparaison du transcriptome des trois tissus de la graine à 12 jap a été réalisée chez la lignée sauvage de référence (Jemalong, A17). Elle a permis de préciser la localisation tissulaire de ces gènes cibles putatifs, de mettre en évidence des voies métaboliques plus spécifiques de l'albumen et de proposer des hypothèses quant à la fonction de ce tissu dans le développement de la graine.

Mots-clés: Développement de la graine, *Medicago truncatula*, caractérisation de mutants, étude comparative du transcriptome, analyse cytologique.

Abstract

In the current context, which necessitates a reduction in inputs in crop systems and boosting of production of plant proteins to reduce France's dependency on feed imports,, growing legumes represents an alternative. Grain legumes are major sources of proteins for animal and human nutrition. In the UMR1347 Agroécologie, the objectives of the study group "déterminismes Génétiques et Environnementaux de l'Adaptation des Plantes à des Systèmes de culture Innovants" (GEAPSI) are to promote legume cultivation and adaptation to environmental stresses, *via* multidisciplinary approaches (genetics, ecophysiology, molecular physiology). This thesis project was carried out in the "Étude des Mécanismes Moléculaires" team, particularly interested in seed quality traits such as protein content or seed size and identification of genes implicated in variations of these traits. Experiments were performed using *Medicago truncatula* as a model species for legumes with a view to transferring the information to the target crop species *Pisum sativum*.

Legume seed size is determined by the embryo's capacity to divide during embryogenesis and to accumulate reserves during seed filling. At early developmental stages, nutrient assimilation occurs predominantly in embryo-surrounding tissues: the endosperm and seed coat. This thesis project aims to evaluate the endosperm contribution to seed development in *M. truncatula*. We have shown several DOF (DNA-binding One Zinc Finger) genes to be expressed in this tissue. They belong to a large family of transcription factors implicated in numerous biological processes, but whose role remain to be elucidated. One of these genes, termed DASH for Dof Affecting Seed embryogenesis and Hormone metabolism, is expressed preferentially in the endosperm during embryogenesis. TILLING and TnT1 mutants isolated for this transcription factor are affected in seed development (abortion at 10 DAP). The cytology of development at early stages (6 to 10 DAP) revealed that the expression of this gene in the endosperm is required for the normal development of the embryo, demonstrating the role of the endosperm in the control of embryogenesis in legumes. A comparative transcriptomics study of *dash* vs wild-type pods allowed us to suggest hypothesis about the function of the *DASH* gene. Evidence for a deregulation of hormone metabolism, in particular for auxin, was obtained, and several potential target genes of this transcription factor were selected. A comparison of the transcriptome of the three tissues of the seed at 12 DAP was carried out for the reference wild-type line (Jemalong A17). This allowed the tissue localization of the target genes, to reveal metabolic pathways preferentially expressed in the endosperm, and to propose hypotheses about the role of this tissue during seed development.

Keywords: seed development, *Medicago truncatula*, mutant characterisation, transcriptomic analyses, cytological studies.

Liste des abréviations

ABA	Acide ABscissique
AIA/IAA	Acide Indole Acétique
alb	albumen
ARF	Auxin Response Factor
BAC	Bacterial Artificial Chromosome
bHLH	basic Helix Loop Helix
BLAST	Basic Local Alignment Search Tool
CGDD	Commissariat Général au Développement Durable
dap	<u>d</u> ays <u>a</u> fter <u>p</u> ollination
DASH	<u>D</u> OF <u>A</u> cting in <u>S</u> eed embryogenesis and <u>H</u> ormone regulation
DOF	DNA binding with One Finger
emb	embryon
EMS	<u>E</u> thyl <u>m</u> ethane <u>s</u> ulfonate
ERF	Ethylene Response Factor
EST	Expressed Sequence Tag
FT	Facteur de Transcription
GES	Gaz à Effet de Serre
jap	jours après polinisation
LEA	Late Embryogenesis Abundant
MtGEA	Medicago truncatula Gene Expression Atlas
NLS	Nuclear Localization Signal
PAC	Politique Agricole Commune
q-PCR	PCR quantitative
QTL	Quantitative Trait Loci
teg	tégument
TAIR	The Arabidopsis Information Resource
TIBA	2,3,5- <u>t</u> riiodo <u>b</u> enzoic <u>a</u> cid
TF	Transcription Factor
TILLING	Targeting Induced Local Lesions IN Genome
UNIP	Union Nationale Interprofessionnelle des plantes riches en Protéines
WT	Wild-type/ génotype sauvage
2-4D	acide 2,4-dichlorophénoxyacétique

Ces travaux de thèse se sont déroulés au sein de l'UMR1347 Agroécologie de Dijon, dans le pôle "déterminismes Génétiques et Environnementaux de l'Adaptation des Plantes à des Systèmes de culture Innovants" (GEAPSI). Je remercie tout d'abord le directeur d'unité Philippe LEMANCEAU et le responsable de pôle Christophe SALON pour m'avoir accueillie au sein du laboratoire. Je tiens également à remercier Richard THOMPSON, mon directeur de thèse de la confiance qu'il m'a accordée pendant ses trois années. J'adresse aussi mes remerciements à ma co-encadrante Karine GALLARDO pour sa patience son attention à mon égard. Je remercie également mes co-encadrantes "non officielles" Christine LE SIGNOR et Vanessa VERNOUD d'avoir été présentes à mes côtés tout au long de ses trois années. Vanessa, je crois que je ne te remercierai jamais assez de ta présence, de ton aide et de ton soutien.

Je souhaite également adresser mes remerciements aux membres du jury Nathalie LEBORGNE-CASTEL, Gwyneth INGRAM, Bertrand DUBREUCQ et Loic RAJOU d'avoir accepté de juger ce travail. Mes remerciements se tournent également vers Peter ROGOWSKY et Bertrand DUBREUCQ qui ont participé aux comités de thèse et avec qui nous avons pu échanger sur les orientations à prendre. Je remercie aussi Julia BUITINK et Sandra PELLETIER de m'avoir initié à l'hybridation de puces et aux dosages de sucres. J'ai passé de très bons moments parmi vous et je remercie toutes les personnes que j'ai pu rencontrer à Angers de m'avoir accueillie au sein de l'UMR IRHS.

Je remercie également les personnes que j'ai pu côtoyer lors des enseignements à l'université de Bourgogne en TP BC1 de biologie cellulaire ou TP BCH4 de biochimie-biologie moléculaire, j'ai beaucoup apprécié de réaliser ces vacations et d'être initié à l'enseignement. Merci de votre confiance.

Enfin je remercie l'Ecole Doctorale E2S, son directeur Thierry RIGAUD et co-directeur Patrick PLESIAT, ainsi que les secrétaires Josette THERY et Corinne AQUILINA, toujours disponibles pour répondre à nos questions. Les bourses de mobilité de l'Ecole Doctorale m'ont permis de financer une partie de ma participation au Model Legume Congress lors de ma 1ère année de thèse et de mon séjour au sein de l'UMR IRHS à Angers.

Un grand merci également à toutes les personnes que j'ai côtoyé et croisé pendant ses trois ans, que ce soit au labo, en serre ou en salle café. Merci à toutes les personnes du pôle GEAPSI de votre accueil et de votre gentillesse! Je remercie particulièrement Karen et Émilie pour leur présence et leurs nombreux coups de mains, merci aux membres de l'équipe EMM : Brigitte, Eric, Nadia, Myriam, Delphine et Catherine. Je remercie aussi Françoise, Greg, Chantal et Catherine...toujours présents lorsque l'on a besoin d'un service. Merci également à Marion et Anne-Sophie pour leur gentillesse et leur sympathie à mon égard.

Un merci tout particulier aux personnes qui se sont succédées dans le bureau: tout d'abord Carlos et Mario, puis Alicia et Arnaud et enfin Anne-Lise et Nadim: merci d'avoir contribué à cette bonne ambiance de travail, merci pour nos échanges et nos fous-rires! Anne-Lise et Nadim: merci de m'avoir supporté ces dernières semaines!

Enfin merci à mes amis et ma famille, de me supporter et me soutenir au quotidien, nombreux sont ceux que j'ai un peu laissé de côté ses derniers temps, mais je ne vous oublie pas et je vous aime!

Sommaire

Introduction

i.1 - Les légumineuses.....	- 2 -
i.1.1 - Une utilisation plus en phase avec l'environnement.....	- 4 -
i.1.2 - Une source de protéine mal exploitée	- 4 -
i.1.3 - Vers une autonomie en protéines : tremplin pour la culture du pois ?.....	- 6 -
i.2 - Etude de la plante modèle <i>Medicago truncatula</i> pour un transfert de connaissances vers l'espèce cultivée <i>Pisum sativum</i>	- 8 -
i.3 – Développement de la graine chez les dicotylédones	- 10 -
i.3.1 – Double fécondation et embryogénèse	- 10 -
i.3.2 - Transition embryogénèse-remplissage et accumulation des réserves.....	- 13 -
i.3.3 – Régulation du développement des tissus de la graine.....	- 15 -
i.4 - Objectifs de la thèse.....	- 16 -

CHAPITRE I - Les facteurs de transcription chez les plantes

I.1 - Introduction.....	- 19 -
I.1.1 - Diversité fonctionnelle des familles de FTs les plus répandues.....	- 21 -
I.1.2 - FTs impliqués dans le développement de la graine.....	- 23 -
I.2 – Focus sur les facteurs de transcription de type DOF (revue).....	- 24 -
I.2.1 - Abstract.....	- 24 -
I.2.2 - Discovery	- 25 -
I.2.3 - Phylogeny of DOF genes.....	- 25 -
I.2.4 - Definition of the DOF domain and its DNA target sites	- 29 -
I.2.5 - Physical and functional association of DOFs with other transcription factors.....	- 30 -
I.2.6 - Relationship of DOF proteins to GATA and steroid hormone receptors	- 32 -
I.2.7 - Classification of DOF gene sequences according to their expression profiles.....	- 32 -
I.2.8 - Identification of conserved domains in DOF protein sequences	- 34 -
I.2.9 - An inventory of biological roles.....	- 38 -

I.2.9.1 - Tissue differentiation	- 38 -
I.2.9.2 - Seed development	- 41 -
I.2.9.3 - Regulation of metabolism	- 43 -
Acknowledgements.....	- 46 -
I.3 - Conclusion	- 46 -
I.4 - Annexes	- 48 -

Chapitre II - Etude d'un facteur de transcription de type DOF impliqué dans le développement de la graine chez *Medicago truncatula*

II.1 - Introduction générale	- 50 -
II.2 - DASH, un régulateur de l'embryogenèse (article en préparation)	- 50 -
II.2.1 - Introduction	- 50 -
II.2.2 - Results.....	- 53 -
II.2.2.1 - <i>DASH</i> encodes an endosperm DOF transcription factor preferentially expressed in endosperm	- 53 -
II.2.2.2 - <i>dash</i> loss-of-function alleles have a seed-lethal or near-lethal phenotype	- 55 -
II.2.2.3 - Insertion of <i>DASH</i> gene in EMS109 complements mutant phenotype	- 57 -
II.2.2.4 - <i>dash</i> mutants are impaired in embryogenesis	- 57 -
II.2.2.5 - A transcriptome comparison of WT and EMS109 mutant pods reveals major transcriptional reprogramming in <i>dash</i> mutants	- 59 -
II.2.2.6 - Down-regulated genes in <i>dash</i>	- 61 -
II.2.2.7 - Up-regulated genes in <i>dash</i>	- 63 -
II.2.2.8 - A striking deregulation of hormone metabolism in <i>dash</i> mutant.....	- 63 -
II.2.2.9 - A link between DASH and auxin signaling.....	- 65 -
II.2.3 - Discussion.....	- 66 -
II.2.3.1 - <i>DASH</i> is expressed in the chalazal endosperm CZE and deregulates a series of genes specifically expressed in the developing endosperm.....	- 67 -
II.2.3.2 - Dash acts partly by repressing expression in the endosperm of genes normally active in other seed/flower tissues	- 68 -
II.2.3.3 - In the absence of DASH, defective endosperm development elicits a generalized stress response	- 68 -

II.2.3.4 - Deregulation of hormone homeostasis in <i>dash</i>	- 69 -
a - Auxin efflux and signaling are impaired in <i>dash</i> seeds.	- 69 -
b - Other hormone signal transduction and metabolism pathways impaired in <i>dash</i> seeds	- 69 -
II.2.4 - Materials and methods	- 70 -
II.2.4.1 - Plant growth conditions and harvest of developing pods	- 70 -
II.2.4.2 - Screening for <i>dash</i> mutant lines.....	- 70 -
II.2.4.3 - Seed trait phenotyping.....	- 71 -
II.2.4.4 - Bioinformatics search for new DOF genes and phylogenetic analysis.....	- 71 -
II.2.4.5 - <i>In situ</i> hybridization	- 71 -
II.2.4.6 - Cell size and cell number measurements.....	- 71 -
II.2.4.7 - Seed clearing.....	- 72 -
II.2.4.8 - Light microscopy	- 72 -
II.2.4.9 - Complementation of <i>dash</i> mutant plants	- 72 -
II.2.4.10 - cDNA synthesis and quantitative real-time PCR analysis.....	- 73 -
II.2.4.11 - Affymetrix Microarray, Data Extraction, and Normalization	- 73 -
II.2.4.12 - Nimblegen Microarray, Data Extraction, and Normalization.....	- 74 -
II.2.4.13 - Phytohormone content measurements and auxin treatments.....	- 75 -
Acknowledgements.....	- 75 -
 II.3 - Données complémentaires à l'article : recherche de cibles potentielles de DASH via une approche transcriptomique.....	- 77 -
II.3.1 – Analyse globale de l'effet génotype quelque soit le stade 8 et 10 jap.....	- 77 -
II.3.2 – Analyse globale plus stringente ($p < 0,01$) des données transcriptomiques à 8 jap	- 79 -
II.3.2.1 - Validation du différentiel d'expression de gènes cibles potentiels de DASH.....	- 81 -
a - Gènes relatifs au métabolisme hormonal	- 84 -
b - Gènes relatifs au métabolisme des sucres	- 84 -
c - Gènes de la classe "26.miscellaneous enzyme families"	- 86 -
d - Gènes de la classe "35.not assigned"	- 88 -
II.3.3 - Recherche des gènes relatifs à la biosynthèse d'AIA.....	- 88 -
II.3.4 - Applications d'auxines exogènes.....	- 92 -
 II.4 – Matériels et méthodes	- 95 -
II.4.1 - Détermination de la teneur en sucre	- 95 -
II.4.2 - Application d'auxine exogène	- 96 -
 II.5 - Conclusion	- 96 -
 II.6 - Annexes	- 98 -

CHAPITRE III - Etude du transcriptome des tissus composant la graine au stade de transition entre l'embryogenèse et le remplissage

III.1 - Introduction	- 107 -
III.2 - Résultats et discussion	- 109 -
III.2.1 - Stratégie d'étude du transcriptome des trois tissus de la graine	- 109 -
III.2.2 - Analyse globale portant sur l'ensemble des gènes exprimés dans les tissus de la graine	- 111 -
III.2.2.1 - Analyse globale des fonctions cellulaires	- 111 -
III.2.2.2 - Analyse des cartes "transcription" générées par le logiciel MapMan	- 113 -
III.2.3 - Analyse ciblée des gènes préférentiellement exprimés dans un tissu	- 115 -
III.2.3.1 - Analyse des spécificités tissulaires	- 115 -
III.2.3.2 - Analyse des classes fonctionnelles	- 117 -
a - Classe fonctionnelle "27.RNA"	- 121 -
b - Classe fonctionnelle "29.protein"	- 123 -
c - Classe fonctionnelle "26.miscellaneous enzyme family"	- 125 -
III.2.4 - Analyse des gènes exprimés pendant l'embryogénèse	- 127 -
III.2.5 - Analyse des gènes impliqués dans le métabolisme de l'AIA	- 129 -
III.3 - Conclusion	- 132 -
III.4 - Matériels et méthodes	- 133 -
III.4.1 - Plan d'expérience	- 133 -
III.4.2 - Obtention du matériel biologique - préparation des échantillons	- 133 -
III.4.3 - Extraction d'ARN et obtention des ARNa (ARN amplifiés)*	- 133 -
III.4.4 - Marquage des ARNa	- 134 -
III.4.5 - Hybridation des puces et scan des lames	- 134 -
III.4.6 - Analyse des données, traitement et conversion	- 135 -
III.5 - Annexes	- 135 -

Conclusion générale et perspectives

Références bibliographiques

Introduction

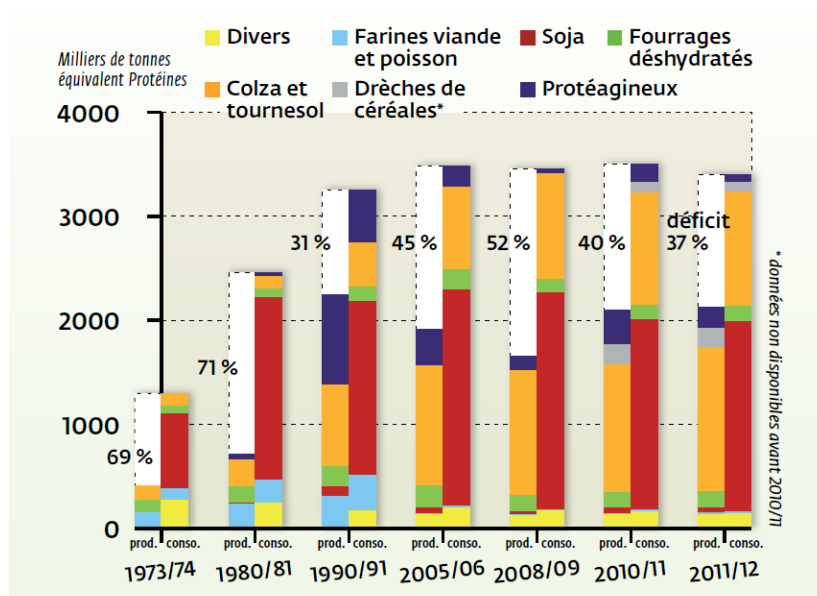


Figure i-1 – Evolution de la production et consommation des matières premières riches en protéines en France mettant en évidence la dépendance protéique. Source : UNIP.

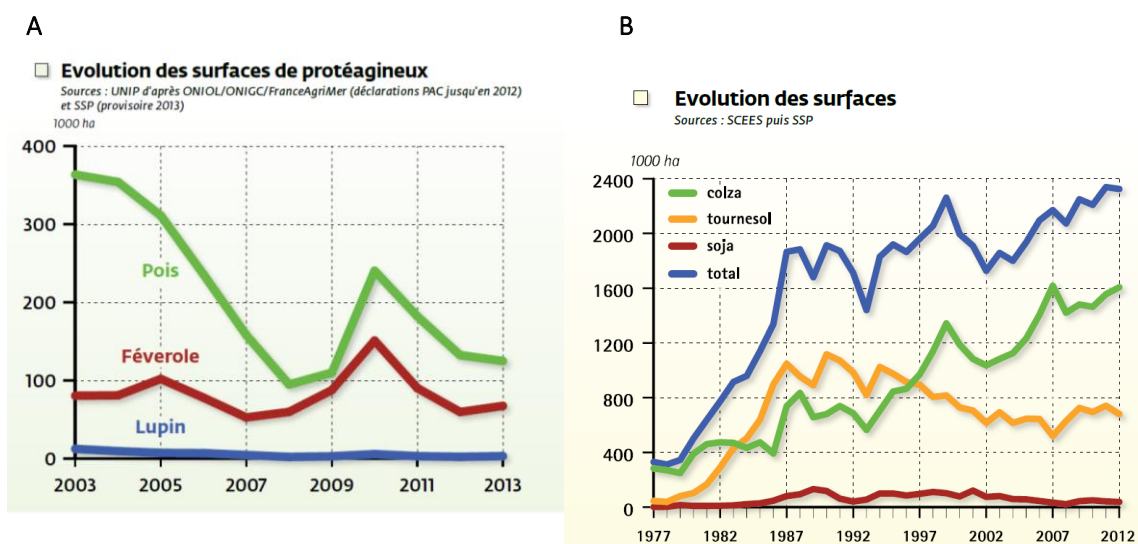


Figure i-2 – Evolution des surfaces cultivées en France. A. Surfaces de protéagineux. B. Surfaces d'oléo-protéagineux (soja) et oléagineux (tournesol et colza). Données ONIDOL d'après des sources collectées auprès de divers organismes. Source : UNIP.

i.1 - Les légumineuses

Le contexte agricole et environnemental actuel tend vers une restriction de l'usage des intrants (engrais, pesticides) dans les systèmes de culture et vers la relance de la production de protéines végétales. La culture des légumineuses constitue une alternative afin de réduire la dépendance protéique des élevages en France vis-à-vis des importations de soja américain ou brésilien. La particularité biologique des légumineuses la plus connue est leur aptitude à s'associer à des bactéries du sol (Rhizobiacées) pour former des organes symbiotiques racinaires ("nodosités" ou "nodules") au sein desquels ces bactéries transforment l'azote atmosphérique en une forme assimilable par la plante. Cette caractéristique leur permet d'assurer leur nutrition sans fertilisation azotée.

On distingue au sein des légumineuses les espèces fourragères (trèfle, luzerne) utilisées directement sous forme de pâturage ou récoltées sous forme de fourrage et les légumineuses cultivées pour leurs graines (et leur contenu en protéines), appelées protéagineux (pois, féverole, lentilles, lupin) destinées à l'alimentation humaine et animale. Enfin les légumineuses telles que le soja et les arachides, produites à la fois pour leur teneur en protéines mais également pour leurs huiles se classent un peu à part, en tant qu'oléo-protéagineux. En Europe, la part de légumineuses cultivées reste faible et insuffisante. Afin de couvrir les besoins en matières premières pour l'alimentation animale, l'Union Européenne est dans l'obligation d'importer pour combler un déficit protéique important de près de 70%, et ce depuis les années 1990 (source UNIP, 2012). Quant à la France, elle est également dépendante de ses importations à hauteur de 40%, un pourcentage certes moins important qu'il y a une trentaine d'année où il s'élevait à plus de 70%, mais qui reste tout de même considérable [figure i-1]. Bien que les importations de tourteaux de soja principalement tendent à se stabiliser ces dernières années, elles se font toujours au détriment de la production de légumineuses qui a fortement chuté [figure i-2A et B]. Il est en effet plus économique d'importer les légumineuses et oléo-protéagineux sous formes de tourteaux que de les produire et les transformer en France. Cette situation est en partie due à des politiques en faveur des cultures de céréales et oléagineux, plus avantageuses économiquement. Aujourd'hui, de nombreux débats tendent à redresser cette situation assez paradoxale en termes d'approvisionnements en protéines végétales et de bilan azoté. Plusieurs rapports du Commissariat Général au Développement Durable (CGDD) font ce constat et apportent de nouvelles réflexions quant à la relance de la culture des légumineuses. La commission "Innovation 2030" proposée au gouvernement en Avril 2013 en vue de définir des axes prioritaires pour le développement économique de la France a établi sept ambitions stratégiques sur lesquelles la France va devoir se distinguer. Ce rapport présenté par Mme Anne Lauvergeon propose parmi ces sept leviers de croissance, une réflexion sur l'avenir des protéines végétales pour une meilleure gestion des ressources. Il s'agit de promouvoir la culture et

Impact		Monétarisation (en milliers d'euros)
Coût pour la balance commerciale agricole	Réduction des exportations de colza	-315 574
	Réduction des exportations de céréales	-340 832
	Réduction des importations de tourteau de soja	429 295
	Sous total : dégradation de la balance commerciale agricole	-227 111
Bénéfices liés à une moindre utilisation d'engrais azotés	Réduction des importations d'engrais azotés	101 116
	Réduction des importations d'énergie fossile	31 137
	Réduction des émissions de GES	56 578
	Sous-total	188 831
Total		-38 280

Figure i-3 – Bilan des coûts et bénéfices liés à la production de légumineuses en France pour remplacer le tourteau de soja importé. Tableau extrait de l'étude du CGDD (2009) Cavaillès et al. 2009.

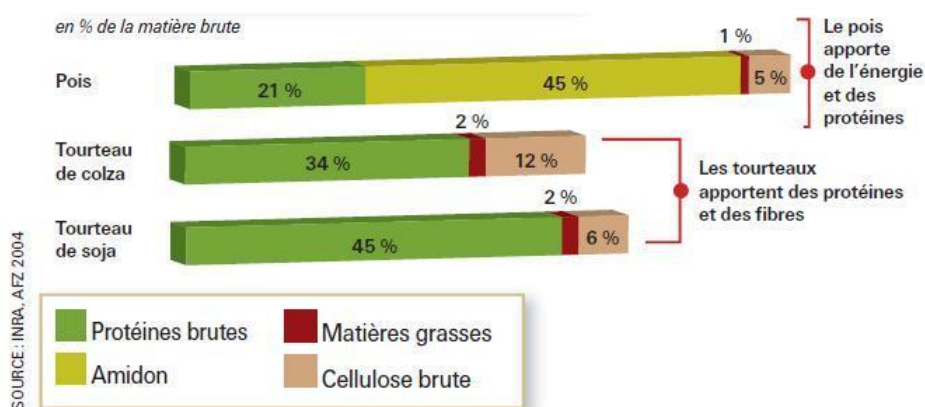


Figure i-4 – Composition nutritionnelle du pois et des tourteaux de colza et soja destinés à l'alimentation animale. Source : Infos Proléa n°71 novembre 2007.

l'utilisation des protéines végétales en alimentation humaine et animale. Au niveau des instituts de recherche, de nombreuses discussions sont également engagées dans le but de valoriser la culture des légumineuses et de déterminer de nouvelles alternatives à leur utilisation (Voisin et al., 2013).

i.1.1 - Une utilisation plus en phase avec l'environnement

Deux atouts majeurs sont mis en avant afin de favoriser l'augmentation de la part des légumineuses au sein des surfaces cultivées : la réduction d'apports azotés et la diminution des émissions de gaz à effet de serre (GES). La faculté de fixation de l'azote atmosphérique que possèdent les légumineuses leur permet de se développer sans fertilisation azotée ; à cela s'ajoute leur capacité à restituer en partie cet azote aux cultures associées ou aux suivantes dans les rotations (Pelzer et al., 2012). Les légumineuses garantissent donc une utilisation moindre d'engrais minéraux azotés et s'imposent comme acteur majeur dans la réorientation vers des systèmes de culture plus respectueux de l'environnement (Nemecek et al., 2008; Siddique et al., 2012). Le plan Ecophyto 2018, établi suite au Grenelle de l'Environnement, traduit cet engagement *"à réduire de 50 % l'usage des pesticides au niveau national dans un délai de dix ans, si possible"* en réduisant notamment *"la dépendance des exploitations agricoles aux produits phytos, tout en maintenant un niveau élevé de production agricole, en quantité et en qualité"*.

Un système de culture plus orienté vers les légumineuses et donc vers un recul de l'utilisation d'engrais azoté permettrait également de réduire l'émission de GES ; l'épandage des engrais azotés provoquant des émissions directes et indirectes de protoxyde d'azote. Une étude du CGDD menée en 2009 a clairement mis en évidence cette réduction des GES en cumulant les économies de GES liées à la réduction de la fabrication et du transport des engrais ainsi que celles liées à un moindre épandage (vaporisation de N_2O) (Cavaillès, 2009). Dans ce même rapport, une étude a été menée afin d'évaluer les conséquences d'un "retour" des légumineuses comme alternative à l'utilisation de tourteaux de soja importés. Un bilan a été dressé afin mettre en évidence les coûts et les bénéfices liés à la production de légumineuses en France (en remplacement des tourteaux de soja importés) et met clairement en évidence l'économie réalisée dans le cadre d'une agriculture plus orientée vers les légumineuses [figure i-3].

i.1.2 - Une source de protéine mal exploitée

Les légumineuses fourragères et les protéagineux constituent une bonne alternative pour combler le déficit protéique et le pois reste l'une des matières premières complémentaires indispensables. Aujourd'hui, il fait l'objet de plusieurs études visant à l'intégrer de façon plus fréquente dans l'alimentation animale, notamment en l'associant aux tourteaux de colza, rendant ainsi cette association compétitive vis-à-vis des tourteaux de soja importés [figure i-4]. En effet, le soja reste

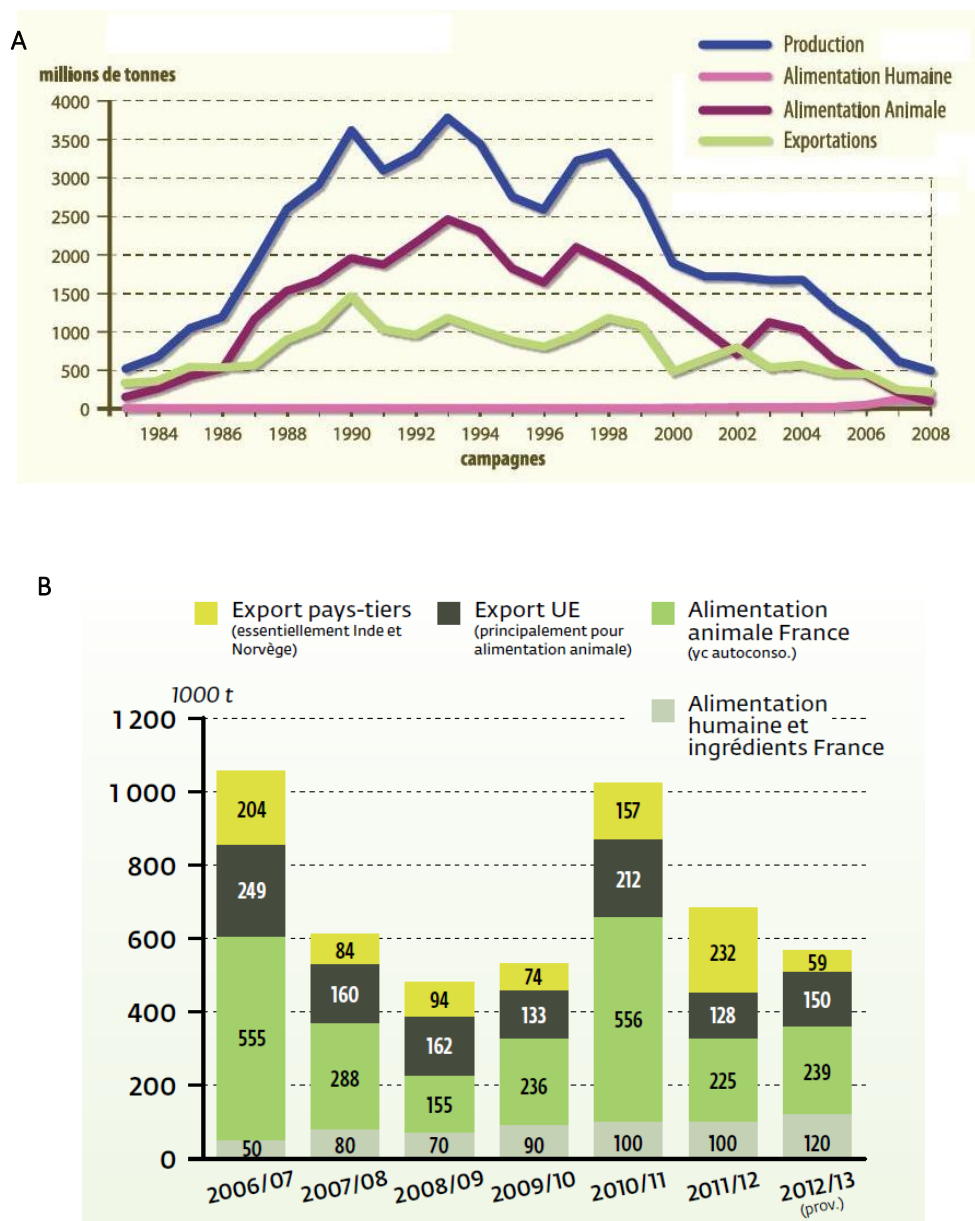


Figure i-5 – Production et principales utilisations du pois protéagineux en France (hors semences). A. Données collectées de 1984 à 2008. B. Données collectées de 2006/2007 à 2012/2013 (prévisionnel). Source : UNIP.

l'élément majoritaire entrant dans la composition des produits d'alimentations animales et est directement importé sous forme de tourteaux ; de gros efforts restent donc à fournir de la production de légumineuses jusqu'à la transformation pour pallier à cette dépendance protéique.

Cette volonté de promouvoir les légumineuses comme source de protéines alternative est identifiable également au niveau politique. Ainsi la loi relative aux organismes génétiquement modifiés, du 25 juin 2008, stipule que le Gouvernement est dans l'obligation d'évaluer *"les possibilités de développement d'un plan de relance de la production de protéines végétales alternatif aux cultures d'organismes génétiquement modifiés afin de garantir l'indépendance alimentaire de la France"*. Un autre texte de loi a également été voté lors du Grenelle de l'Environnement, dont l'objectif est *"de réduire la dépendance des systèmes de production animale aux matières premières importées entrant dans la composition des produits d'alimentation animale, notamment en relançant la production des cultures de protéagineux et autres légumineuses"*.

i.1.3 - Vers une autonomie en protéines : tremplin pour la culture du pois ?

La relance des légumineuses apparaît comme une solution envisageable pour contribuer à l'indépendance protéique de la France. La culture du pois se retrouve donc au centre des débats de relance des protéagineux à destination de l'alimentation animale. L'évolution de la production et de l'utilisation du pois protéagineux indique que cette culture a subi les revers des politiques agricoles françaises (et européenne), plus orientées en faveur des céréales sur cette période [figure i-5A et B]. Pendant les années 1990, le pois était beaucoup utilisé par les fabricants d'aliments pour le bétail, du fait de sa teneur importante en protéine d'où sa forte production et consommation. C'est ensuite l'impact des réformes de la PAC des années 2000 qui ont conduit à cette diminution de l'utilisation, et par conséquent à la baisse de la production [figure i-5A]. En 2009, face à ces résultats décourageants, le ministère de l'Agriculture a attribué 40 millions d'euros d'aides à la relance des cultures de protéagineux avec un objectif de 400 000 hectares de protéagineux en 2012 contre 165 000 hectares en 2008 (données Réussir grandes cultures n°224). Cette relance a effectivement eu lieu en 2010/2011 où l'on observe une forte production et consommation *via* l'alimentation animale, mais dès 2011/2012, la production et par conséquent la consommation ont à nouveau fortement chuté [figure i-5B].

De nombreuses interrogations subsistent encore aujourd'hui quant à l'avenir du pois au sein de notre agriculture, et la question du devenir des protéagineux reste sans réponse, en témoigne Patrice Rialland responsable des achats du groupe Sanders et administrateur de l'UNIP : *"Le pois lutte à l'amont pour conserver sa place dans l'assolement, dans une vive concurrence avec l'ensemble des grandes cultures qui bénéficient d'une visibilité de prix et d'un marché structuré. En aval, le pois est confronté au blé pour l'énergie et aux tourteaux de colza et de soja, très disponibles, au sein d'un portefeuille de matières premières bien organisé avec une visibilité de long terme sur les prix. "*

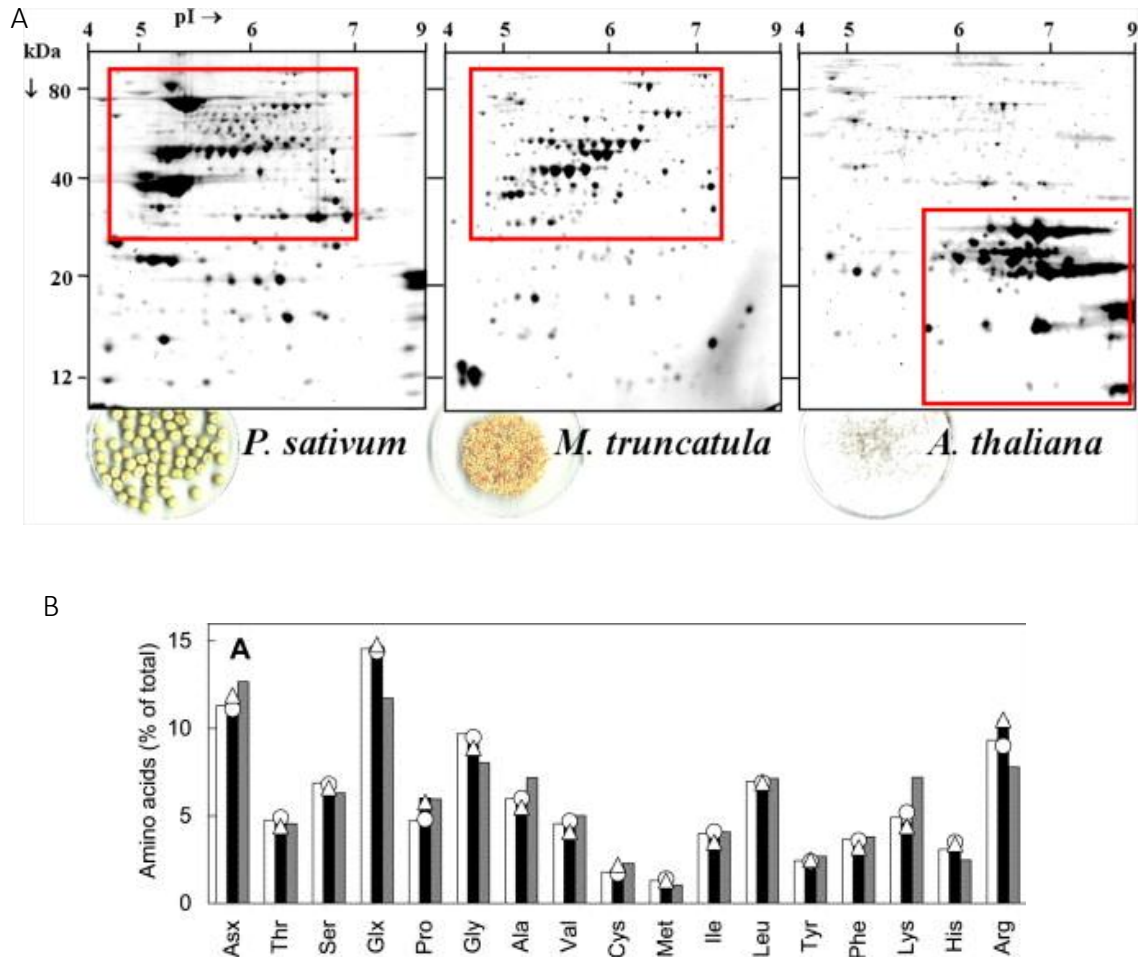


Figure i-6 – A : Comparaison du protéome des graines matures de pois, *M. truncatula* et *A. thaliana*. Les images correspondent aux gels 2-D d'extraits de protéines totales solubles (200µg) séparées suivant un gradient de pH (3 à 10) puis en fonction de la taille par un gel SDS-PAGE, coloration au bleu de Coomassie G-250. Figure extraite de Gallardo et al. 2008.

B : Composition en acides aminés des graines matures de huit génotypes chez *M. truncatula* (en blanc et noir ; lieux de cultures indépendants) et un génotype de pois (en gris). Figure extraite de l'étude de Djemel et al. 2005.

Les objectifs de recherche sur le pois visent donc une meilleure connaissance et structuration de cette culture dans le but de concevoir un système compétitif et durable, en tenant compte des variations environnementales. Les préoccupations actuelles portent sur la stabilité de la production en termes de tolérance aux stress (changements climatiques et résistance aux maladies) et de maîtrise des rendements et de la qualité des graines (taille de la graine, teneur en protéines). Des études sont en cours afin de mieux connaître et caractériser la tolérance du pois vis-à-vis de ces différents stress biotiques et abiotiques dans différents instituts de recherche, dont l'INRA, engagé auprès de l'UNIP et PROLEA pour le maintien de la filière protéagineux. D'autres axes de recherche visent des caractères plus qualitatifs et/ou quantitatifs tels que l'augmentation ou le maintien du rendement en graines ou encore l'amélioration de la qualité et de la teneur en protéines des graines. Un projet Investissements d'avenir ciblé sur le pois (PeaMUST) a démarré en 2012. Porté par J. Burstin (UMR Agroécologie Dijon), ce projet a pour objectif de développer de nouvelles variétés de pois et d'optimiser leurs interactions pour stabiliser le rendement et la qualité des graines de pois, dans le contexte du changement climatique et de la réduction de l'utilisation des pesticides.

*i-2 - Etude de la plante modèle *Medicago truncatula* pour un transfert de connaissances vers l'espèce cultivée *Pisum sativum**

Afin d'étudier les mécanismes moléculaires impliqués dans le développement de la graine, l'espèce *Medicago truncatula* (*M. truncatula*) est utilisée comme modèle des légumineuses. *M. truncatula* possède un génome de 450Mb, sept fois plus petit que celui du pois (Barker et al., 1990). C'est une plante diploïde ($2n = 16$) caractérisée par un mode de reproduction autogame (Rose, 2008) et une croissance indéterminée permettant d'obtenir au sein d'une même plante de nombreuses gousses à différents stades de maturité physiologique (Moreau et al., 2006), une caractéristique intéressante pour les projets axés sur le développement de la graine. *M. truncatula* est largement utilisée dans les laboratoires de recherche et ce depuis de nombreuses années (Barker et al., 1990; Frugoli and Harris, 2001). Cette plante est une espèce sauvage annuelle du pourtour méditerranéen appartenant à la famille botanique des Fabacées, plus couramment appelées légumineuses. Due à son génome de petite taille, elle constitue une alternative à l'étude des mécanismes moléculaires, même si la plante modèle des crucifères, *Arabidopsis thaliana* (*A. thaliana*), a permis des avancées considérables dans la compréhension des bases génétiques et moléculaires du fonctionnement des plantes. Une forte syntenie est observée entre *M. truncatula* et *Pisum sativum* (*P. sativum*) et d'autres espèces de la famille des Fabacées comme le soja (*Glycine max*) (Choi et al., 2004), ce qui permet un transfert de connaissances entre espèces. La composition protéique des graines matures de *M. truncatula* et *P. sativum* est très similaire (Gallardo et al., 2008) [figure i-6A] et des travaux menés au laboratoire ont

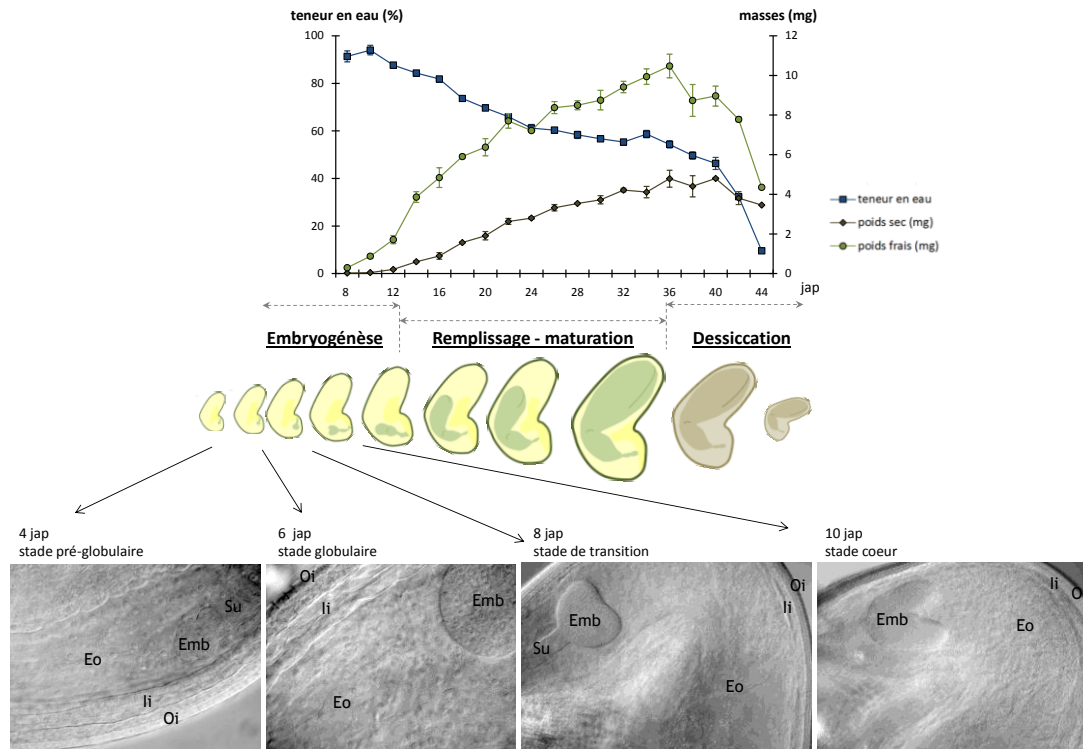


Figure i-7 – Etapes du développement de la graine chez *M. truncatula*. Graphe présentant les poids frais et poids secs ainsi que la teneur en eau des graines de 8 à 44 jap (jours après pollinisation), extrait de Gallardo et al. 2003. Observations cytologiques (réalisées lors de la thèse) du développement de la graine durant les stades d’embryogénèse précoce, du stade pré-globulaire à 4 jap au stade cœur à 10 jap. Emb, embryon; Eo, albumen; li, tégument interne; Oi, tégument externe; Su, suspenseur.

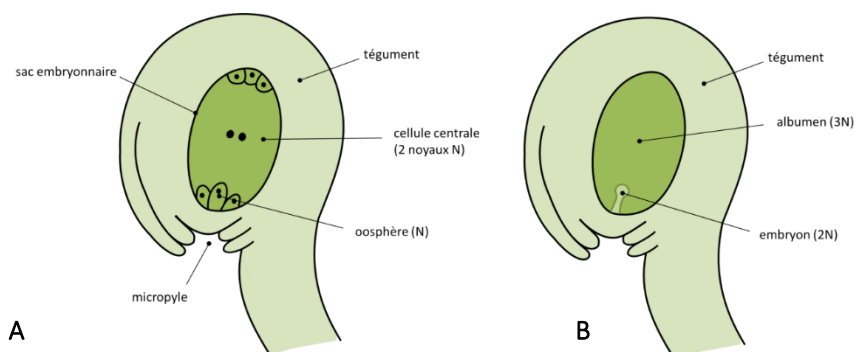


Figure i-8 – Représentation de la formation de la graine avant et après la double fécondation. A. Gamète femelle composée par le sac embryonnaire protégé par le tégument. Le micropyle permet aux gamètes mâles d’atteindre le sac embryonnaire afin de féconder l’oosphère et les noyaux de la cellule centrale. B. Les trois tissus de la graine, dont l’embryon et l’albumen issus de la double fécondation. Figure reproduite d’après Reiser et al. 1993.

révélé une conservation entre ces deux espèces des mécanismes de régulation de la taille (D'Erfurth et al. 2012) et de la composition protéique des graines (données non publiées). On note également une forte similitude entre ces deux espèces de la composition en acides aminés des graines matures (Djemel et al., 2005); celles-ci étant pauvres en acides aminés soufrés et riches en lysine [figure i-6B].

Les ressources en matière de données génomiques (séquençage, carte génétique, banques de BACs et d'ESTs, puces..) se sont multipliées afin de mieux appréhender les recherches sur les spécificités des légumineuses (Frugoli and Harris, 2001; Rose, 2008; Young et al., 2011). L'étude de ces caractéristiques propres aux légumineuses, telles que les différents processus biologiques et physiologiques liés au développement de la plante (capacité de fixation de l'azote, architecture racinaire associée à la nodulation, etc..) et des graines (embryogenèse et remplissage), ou encore aux capacités d'adaptation face à des stress biotiques et abiotiques, est nécessaire en vue de promouvoir la culture de légumineuses et de valoriser leurs produits. Les recherches des mécanismes régulant la composition et la taille de la graine permettront par exemple d'identifier des variétés agronomiques répondant à des critères de qualité de la graine (ex. teneur en protéine stable et élevée).

i.3 - Développement de la graine chez les dicotylédones

La qualité de la graine dépend à la fois du génotype et des facteurs environnementaux intervenant lors des différentes phases de développement, et notamment lors de l'embryogenèse, une étape-clé dans la détermination du nombre de cellules de la graine et donc de sa capacité à accumuler des réserves. L'embryogenèse constitue la première phase du développement de la graine. Chez *M. truncatula*, elle s'arrête à 12 jours après pollinisation (jap), stade de transition vers le remplissage. La graine accumule des réserves et des protéines de maturation de type LEA (Late Embryogenesis Abundant Proteins) jusqu'à environ 36 jap, stade à partir duquel elle se déshydrate rapidement pour ne conserver que 10% d'eau à maturité [figure i-7].

i.3.1 – Double fécondation et embryogenèse

Au cours de son développement, la graine passe d'un contrôle maternel à un contrôle zygotique mené par l'embryon. Alors que le tégument dérive directement des tissus maternels, l'embryon et l'albumen sont quant à eux issus de la double fécondation, qui fait intervenir deux gamètes mâles : l'un va féconder l'oosphère et le second va fusionner avec les noyaux polaires de la cellule centrale [figure i-8A]. La première fécondation évolue ensuite en un zygote diploïde qui deviendra l'embryon (2N) ; la deuxième conduira à la formation d'un zygote triploïde, à la base de l'albumen [figure i-8B]. La fécondation de l'oosphère et la triple fusion avec les noyaux polaires sont l'une et l'autre nécessaires à la

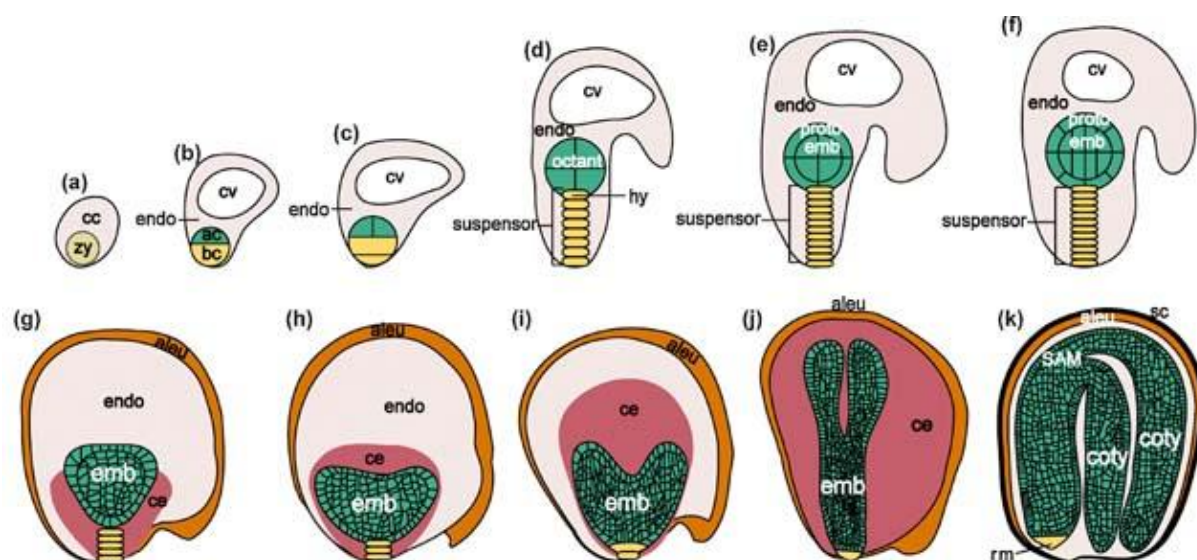


Figure i-9 - Développement de la graine chez *A. thaliana*. (a) Zygote (zy) et cellule centrale (cc) suite à la première fécondation. (b) Zygote ayant subi une première division qui aboutit à la formation d'une cellule apicale (ac) et d'une cellule basale (bc) ; albumen (endo) issu de la seconde fécondation avec les noyaux de la cellule centrale ; vacuole centrale (cv). (c) Evolution des divisions de l'albumen et de l'embryon au stade quatre cellules. (d) Embryon au stade huit cellules (octant) et évolution du suspenseur à partir des cellules basales, la cellule à la base de l'embryon à la jonction avec le suspenseur est qualifiée d'hypophyse (hy). (e) Embryon au stade 16 cellules, la couche de cellules la plus externe se différencie en protoderme. (f) Embryon au stade globulaire. (g) Embryon au stade transition (approximativement 200 cellules), début de cellularisation de l'albumen dans la région bordant l'embryon (ce) et dans la région bordant le tégument (aleurone: aleu). (h) Embryon au stade cœur, les cotylédons (coty) commencent à se différencier et la cellularisation de l'albumen progresse. (i) Embryon au stade torpille, le suspenseur commence à se dégrader. (j) Embryon au stade cotylédons linéaires, l'albumen cellularisé occupe tout l'espace entre l'embryon et le tégument (sc). (k) Au fur et à mesure que l'embryon se développe, l'albumen est dégradé (mort cellulaire programmée), le méristème racinaire (rm) et le méristème apical (SAM) se distinguent.

Figure extraite de Agarwal et al. 2011

formation d'une graine viable (Reiser and Fischer, 1993; Goldberg et al., 1994). Les téguments de l'ovule deviennent le tégument de la graine. Ce dernier joue un rôle important dans le développement de l'embryon en le protégeant et en assurant le transfert de nutriments. En effet, aux stades précoces du développement, l'assimilation de nutriments est réalisée majoritairement par les tissus entourant l'embryon : l'albumen et le tégument. Une accumulation temporaire de réserves a lieu dans ces tissus, ainsi qu'un métabolisme intermédiaire intensif servant à convertir et fournir des nutriments pour la croissance de l'embryon.

Une étude de l'embryogénèse et du développement de l'embryon chez *M. truncatula* menée par Wang et al. 2011 a permis d'observer une organisation structurale des noyaux au sein du sac embryonnaire similaire à celle présentée dans la figure i-8 (Reiser and Fischer, 1993). Cette étude a également permis de mettre en évidence l'importance du suspenseur pour le développement de l'embryon, et la distribution des corps protéiques et corps gras dans l'embryon (Wang et al., 2011). L'embryogénèse se caractérise par de nombreuses divisions cellulaires dans l'embryon (Abirached-Darmency et al., 2005; Wang and Grusak, 2005). La figure i-9, extraite de Agarwal et al. 2011, présente les premières étapes du développement de l'embryon chez *A. thaliana* et met en évidence les nombreuses divisions cellulaires ayant lieu dans ce tissu au cours de l'embryogénèse (Agarwal et al., 2011). Le zygote issu de la première fécondation se divise en deux cellules: une cellule basale qui sera à l'origine du suspenseur et une cellule apicale qui se différenciera en embryon [figures i-9a à f]. L'albumen syncytial commence à se cellulariser autour de l'embryon lorsque celui-ci a atteint le stade transition et autour du tégument, formant une couche cellulaire appelée aleurone [figure i-9g]. L'albumen continue ainsi à se cellulariser jusqu'à occuper tout l'espace entre l'embryon et le tégument au stade cotylédons linéaires [figure i-9j]. C'est à partir de ce stade que commence la dégradation progressive de l'albumen au profit du remplissage de l'embryon jusqu'au stade mature [figure i-9k].

La caractérisation du développement de la gousse et du tégument décrite par Wang et Grusak (Wang and Grusak, 2005), qui souligne l'importance des faisceaux vasculaires dans le développement du tégument, apporte quelques indications quant au développement de l'albumen. Ainsi, les noyaux de l'albumen sont visibles à l'intérieur du sac embryonnaire dès le stade globulaire, et les couches de l'albumen sont observées autour du sac embryonnaire lorsque l'embryon est au stade cœur. Des observations cytologiques réalisées au laboratoire chez *M. truncatula* [figure i-7] complètent ces informations. Ainsi, les noyaux de l'albumen sont également visibles au stade pré-globulaire et l'albumen syncytial semble se développer en albumen cellularisé dès le stade globulaire, le front entre cet albumen syncytial et l'albumen cellularisé se distingue dès 8 jap. Ces observations sont similaires à celle observées chez *A. thaliana* où la cellularisation de l'albumen est d'abord marquée dans les régions bordant l'embryon [figure i-9].

Une étude portant sur le développement de la graine pendant les stades précoces de l'embryogénèse chez *M. truncatula* présente les différentes étapes du développement de la fleur à la gousse jusqu'au stade 8 jap, ainsi que le développement de l'embryon du stade globulaire au stade torpille (Kurdyukov et al., 2014). Ces travaux ont mis en évidence le profil d'expression de plusieurs FTs au cours de ces stades d'embryogénèse précoce, soulignant ainsi des spécificités fonctionnelles.

i.3.2 - Transition embryogénèse-remplissage et accumulation des réserves

Comme énoncé précédemment, l'embryogénèse est caractérisée par de nombreuses divisions cellulaires, le rythme de ces divisions diminue ensuite fortement lors de la transition vers le stade de remplissage/maturation où une forte activité de synthèse des protéines de réserve (légumine; viciline) puis de maturation (protéine de type LEA) est observée (Gallardo et al., 2007). Une étude comparative du transcriptome et protéome de la graine au cours du développement a mis en évidence une compartimentation des activités métaboliques entre les différents tissus de la graine mais également une coordination importante entre l'accumulation des transcrits et des protéines de réserve et de maturation (Gallardo et al., 2007). Cette analyse a permis de constituer des cartes de régulation fonctionnelles relatives au métabolisme des acides aminés soufrés qui entrent dans la composition des protéines de réserve, ou encore une vue plus globale de la compartimentation tissulaire des gènes et intermédiaires réactionnels impliqués dans la biosynthèse des acides aminés au début du remplissage de la graine (Gallardo et al., 2007). D'autres approches plus ciblées sur les facteurs de régulation ont été utilisées. L'étude du protéome nucléaire à 12 jap, stade de transition entre l'embryogénèse et le remplissage (Repetto et al., 2008), a mis en évidence plusieurs protéines susceptibles de réguler l'activité transcriptionnelle des graines *via* des remodelages au niveau de la chromatine. Le faible nombre de facteurs de transcription identifiés par cette approche a conduit le laboratoire à étudier, *via* une approche de PCR quantitative haut débit, les profils d'expression de 700 facteurs de transcription lors des derniers stades d'embryogénèse jusqu'à dessiccation de la graine (Verdier et al., 2008). Plusieurs facteurs de transcription spécifiques des différentes phases du développement de la graine ont été identifiés et des collections de mutants TILLING développées (Le Signor et al., 2009) afin de valider la fonction de ces gènes. Plusieurs gènes candidats sélectionnés sont fortement exprimés au stade de transition entre l'embryogénèse, caractérisée par de nombreuses divisions cellulaires, et le remplissage.

Ce stade de transition embryogénèse-remplissage est déterminant pour le contrôle du nombre de cellules puisqu'il marque le ralentissement des divisions cellulaires au sein de l'embryon et le début de l'accumulation des réserves de la graine, associée à l'expansion des cellules de l'embryon. L'importance de la composition en sucres a été soulignée dans cette phase de transition, caractérisée par une forte teneur en sucre C12 dans l'embryon (Djemel et al., 2005). Les sucres tels que le saccharose sont nécessaires afin d'induire la différenciation cellulaire, alors que les hexoses sont plutôt impliqués

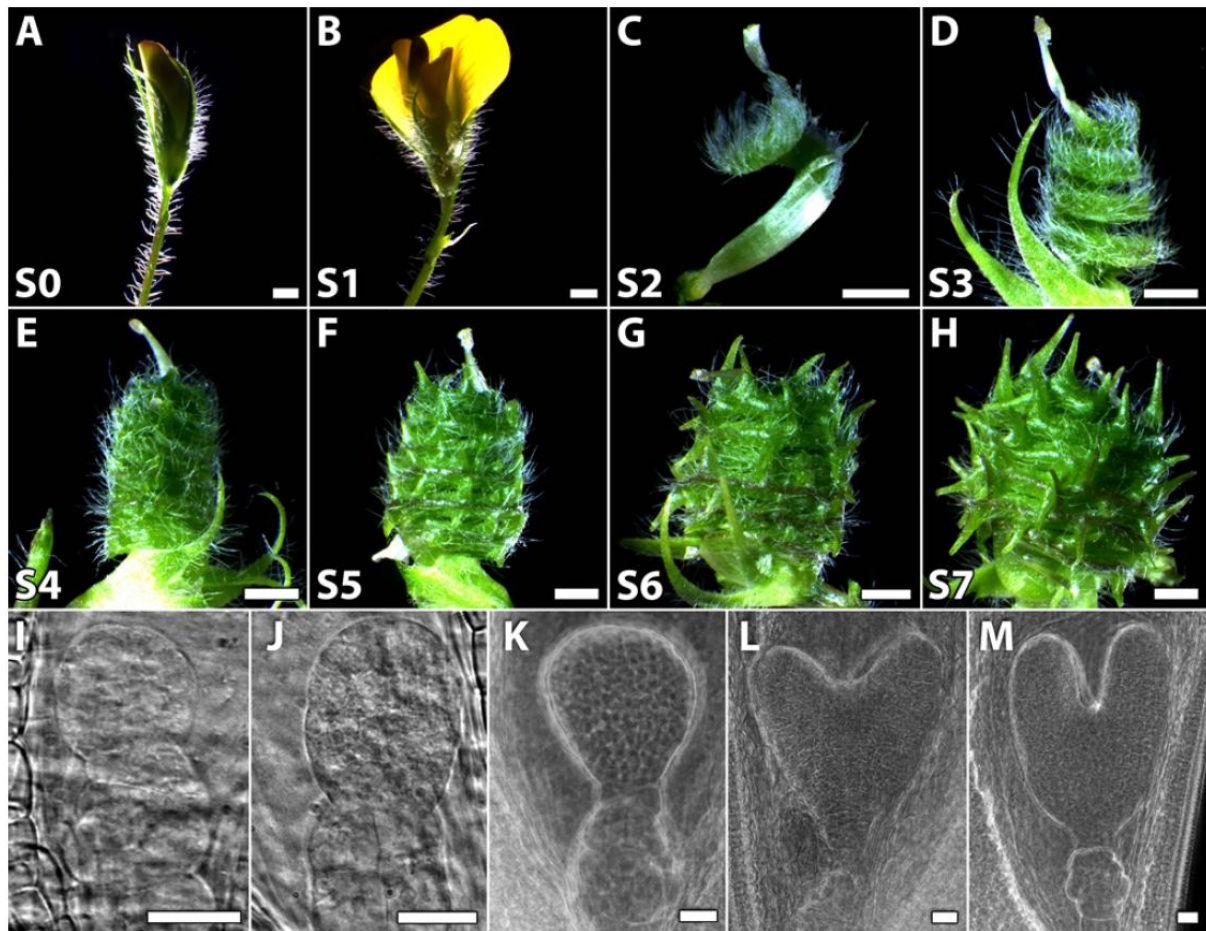


Figure i-10 – A à H : Développement de la fleur à la gousse au cours des différents stades de l’embryogénèse. A : fleurs deux jours avant l’anthèse. B : fleurs ouvertes (à l’anthèse). C à H : gousse en développement (stade S2 : 2,5-3 jap ; stade S3 : 3-4 jap, stade S4 : 3,5-4,5 jap ; stade S5 : 4-5,5 jap ; stade S6 : 5-6,5 jap et stade S7 : 6-7,5 jap). Echelle 2mm.

I à M : Développement de l’embryon des stades S3 à S7. I : stade 8-16 cellules (S3) ; J : stade pré-globulaire (S4) ; K : stade globulaire (S5) ; L : stade cœur (S6) ; M : stade torpille (S7). Echelle 25µm

Figure extraite de l’article de Kurdyukov et al. 2014.

dans le contrôle des divisions cellulaires. Les auteurs ont également montré que le stachyose est le principal oligosaccharide présent dans les graines matures de *M. truncatula*. La phase de remplissage/maturation est caractérisée par l'accumulation dans l'embryon des réserves indispensables à la survie de la graine. La composition en lipides, protéines et amidon, de même que les signaux hormonaux sont en effet déterminants pour la germination future de la graine (Rajjou et al., 2012). En effet, c'est également pendant cette phase de remplissage/maturation que la graine acquiert sa vigueur germinative (Gallardo et al., 2003 ; Buitink et al., 2006; Vandecasteele et al., 2011) ainsi que sa tolérance à la dessiccation (Chatelain et al., 2012). Bien que les conditions environnementales jouent un rôle non négligeable dans le processus de germination, la graine doit, pendant son développement préparer sa dormance et sa germination future et mettre en place toute une régulation afin d'acquérir ce pouvoir germinatif (Holdsworth et al., 2008). Enfin, lorsque la graine atteint sa maturité physiologique, ses activités métaboliques sont réduites et elle se déshydrate au cours de la phase de dessiccation (Chatelain et al., 2012), l'abscission des gousses a lieu vers 44 jap.

i.3.3 – Régulation du développement des tissus de la graine

Au cours du développement de la graine, les trois tissus évoluent conjointement avec une coordination plus marquée entre l'albumen et l'embryon ; l'albumen constituant un tissu de réserve pour couvrir les besoins de l'embryon (Dumas and Rogowsky, 2008). Chez les céréales ce tissu nourricier persiste et sert de réserve pour la germination future de l'embryon (graines albuminées). Chez la plupart des légumineuses, ce tissu est transitoire et est utilisé par l'embryon au cours de son développement, les composés de réserves sont alors stockés dans les cotylédons, on parle de graines exalbuminées. Chez *A.thaliana*, le développement de l'albumen et notamment sa cellularisation influe fortement sur celui de l'embryon (Hehenberger et al., 2012). En effet, l'étude de mutants *fertilization independent seed 2 (fis2)* (Chaudhury et al., 1997) et *endosperm defective 1 (ede1)* (Pignocchi et al., 2009) déficients pour le contrôle des divisions cellulaires de l'albumen montre un arrêt précoce du développement de l'embryon ou un retard de développement de l'embryon associé à un retard de cellularisation de l'albumen par rapport aux observations dans les graines de type sauvage. La cellularisation de l'albumen serait associée à une réduction de la teneur en sucre C6; le transport de ces sucres vers l'embryon serait altéré chez les mutants, conduisant ainsi à une accumulation des sucres C6 dans l'albumen (Hehenberger et al., 2012). D'autres paramètres sont impliqués dans le développement conjoint de l'albumen et de l'embryon. Ainsi, un peptide signal CLE8 (CLAVATA3/EMBRYO SURROUNDING REGION-RELATED8), produit par l'albumen et l'embryon lors de l'embryogénèse précoce a été mis en évidence dans le contrôle des signaux régulant la prolifération de l'albumen et les divisions cellulaires de l'embryon (Fiume and Fletcher, 2012). Un facteur de transcription (FT) de type bHLH a également été décrit comme impliqué dans le développement de l'embryon (Kondou et al., 2008; Yang et al., 2008). Ce FT ZHOUP1 (ZOU) ou

RETARDED GROWTH OF EMBRYO (RGE) est spécifiquement exprimé dans les régions de l'albumen bordant l'embryon et les graines produites par les mutants d'insertion T-DNA pour ce gène présentent des défauts de développement de l'embryon qui seraient liés à la structure de l'albumen. En effet chez ces mutants, l'albumen persiste tout au long du développement de la graine, empêchant ainsi un développement normal de l'embryon, et conduisant à des graines asséchées de petite taille. Les auteurs ont également montré que ce FT est impliqué dans la régulation du gène *ALE1* (*abnormal leaf shape1*) dont la mutation (transposon *Ac* transposase) induit également des défauts dans la morphologie structurelle de l'embryon et dans la formation de la cuticule à sa surface (Tanaka et al., 2001; Yang et al., 2008). Ce défaut de formation de la cuticule empêche l'adhésion des cellules de l'albumen et conduit à la fusion des cotylédons. Les observations conduites dans ces travaux indiquent une perte excessive de la teneur en eau des graines, phénotype observé à la fois dans les graines des mutants *zou* et *ale*. La communication entre l'albumen et l'embryon est donc déterminante pour un développement normal de l'embryon. Plusieurs interactions ont été recensées dans une étude des signaux contrôlant le développement de la graine chez *A. thaliana*. Des régulateurs du développement et de la cellularisation de l'albumen tels que *AP2* (*APETALA2*) ou *HAIKU* (*IKU*) et *MINISEED* (*MINI3*) influencent fortement le développement de l'embryon (Sun et al., 2010). On retrouve ainsi de nombreux facteurs de transcription impliqués dans la régulation du développement des tissus et l'étude des voies métaboliques conduisant au développement de la graine semble complexe (Xiang et al., 2011). Les changements intervenant pendant les phases de développement de la graine sont nombreux et leur régulation nécessite une importante coordination.

i.4 - Objectifs de la thèse

La contribution de l'albumen au développement de la graine chez *M. truncatula* a été étudiée lors de la thèse en utilisant deux approches : (i) une approche ciblée sur un FT de type DOF, préférentiellement exprimé dans ce tissu à des stades précoces du développement de la graine et pour lequel des mutants *TILLING* et *Tnt1* étaient disponibles, et (ii) une approche globale (transcriptomique) pour explorer les voies métaboliques actives dans l'albumen au stade de transition entre l'embryogenèse et le remplissage.

Le premier chapitre de la thèse présente le rôle des FTs de type DOF (DNA binding with One Finger) dans le développement des plantes et des graines. Ce chapitre a fait l'objet d'une revue: "The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants" (Noguero et al., 2013).

La seconde partie concerne l'étude d'un FT de type DOF précédemment identifié par une approche transcriptomique (Verdier et al., 2008). Ce chapitre présente le manuscrit en relecture "DASH transcription factor regulates seed size and embryo morphogenesis by disturbing auxin homeostasis" et

des travaux complémentaires réalisés lors de la thèse. Il présente l'étude de mutants pour ce FT et propose des hypothèses quant au rôle de ce FT grâce à des descriptions phénotypiques, cytologiques et transcriptomiques. Dans ce chapitre sont aussi discutés les cibles possibles de *DASH* et le rôle de l'auxine dans le contrôle du développement des graines.

Enfin, le troisième chapitre présente l'étude du transcriptome des trois tissus de la graine à un moment clé du développement : la transition entre l'embryogénèse et le remplissage. Cette étude relate les différentes voies métaboliques et la répartition des gènes par classe fonctionnelle dans les trois compartiments de la graine (tégument, albumen et embryon), mettant en évidence des spécificités métaboliques tissulaires et soulignant le rôle clé de l'albumen dans le contrôle du développement de la graine.

CHAPITRE I

-

Les facteurs de transcription chez les plantes

1.1 - Introduction

Les facteurs de transcription (FTs) sont des éléments clés dans la régulation fonctionnelle et le développement des organismes vivants. Chez les plantes, les FTs constituent une piste d'étude importante pour la compréhension des voies métaboliques et des mécanismes fonctionnels contrôlant les différents traits phénotypiques déterminants utilisés en sélection variétale (Arce et al., 2008). Ils ont effectivement un rôle primordial dans le développement puisqu'ils interviennent sur l'ensemble des voies de régulation. En effet, toutes les fonctions biologiques nécessitent un contrôle par les FTs, qui régulent l'expression des gènes i) dans les différents compartiments de la plante, ii) à des moments clés du développement, iii) en réponse à des stimuli externes. Ils font aujourd'hui l'objet de nombreuses études dans le but de mieux comprendre leur structure et leur mode de fonctionnement. Les FTs sont capables de reconnaître un motif sur la région promotrice de leurs gènes cibles, et de s'y lier de façon spécifique afin d'activer ou de réprimer la transcription. Quelquefois, l'activation ou la répression de cette transcription nécessite la présence simultanée de plusieurs FTs: ils ne se fixent alors pas directement à l'ADN mais interagissent entre eux (oligomérisation). Chez les plantes, les TFs sont généralement constitués de quatre types de domaines : le domaine de liaison à l'ADN, un domaine de régulation de la transcription également appelé domaine de transactivation (ou de répression), le site d'oligomérisation et le domaine de localisation nucléaire. Les TFs sont ainsi capables de reconnaître et de lier spécifiquement des éléments cis-régulateurs grâce à leur domaine de liaison à l'ADN afin de réguler la transcription de gènes cibles *via* leur domaine de régulation de la transcription. Le domaine d'oligomérisation permet quant à lui l'association à d'autres FTs afin de favoriser la liaison au motif reconnu dans la séquence promotrice dans le but d'amplifier la réaction d'inhibition ou d'activation de la transcription. Le signal de localisation nucléaire (NLS), qui se compose de séquences variables selon les FTs et présentes en une ou plusieurs copies, conduit à l'entrée des FTs dans le noyau des cellules. Ces quatre domaines sont communs à presque tous les FTs, avec de fortes similarités entre les séquences, ce qui laisse supposer l'existence d'un ancêtre commun (Purugganan and Wessler, 1994) ; les mutations dans les régions conservées ayant conduit à l'évolution et à la spécificité des domaines de liaison à l'ADN (Liu et al., 1999 ; Liu et al., 2001). En fonction du gène cible, les FTs peuvent réguler positivement ou négativement la transcription, de même qu'ils peuvent jouer ce double rôle d'activateur ou de represseur dans certains cas. Ainsi, le FT *Viviparous-1* (VP1) peut sous certaines conditions activer la transcription du gène C1 qui contrôle la pigmentation des grains de maïs (Coe et al., 1988). Dans un autre contexte fonctionnel, VP1 peut devenir un represseur de l'expression de l' α -amylase et ainsi réguler négativement la germination (Hoecker et al., 1995; Paek et al., 1998; Hoecker et al., 1999). Il existe donc pour un même FT plusieurs domaines qui contrôlent l'activation ou la répression de la

CHAPITRE -1-

type de TF	description	Arabidopsis	Maïs	Riz	Medicago	Soja
bHLH	<u>b</u> asic <u>h</u> elix <u>l</u> oop <u>h</u> elix	225	308	211	133	480
MYB	du gène v-myb issu du virus du <u>m</u> yeloblastome aviaire	168	203	130	101	369
ERF	<u>e</u> thylene <u>r</u> esponse <u>f</u> actor	139	205	163	106	330
NAC	<u>N</u> AM (non apical meristem), <u>A</u> TAF (Arabidopsis transcription activation factor) and <u>C</u> UC (cup shaped cotyledon)	138	190	170	75	247
C2H2	X2- <u>C</u> ys-X2- <u>C</u> ys-X9- <u>H</u> is-X3- <u>H</u> is-X2	116	179	135	112	267
MYB_related	du gène v-myb issu du virus du <u>m</u> yeloblastome aviaire	97	169	106	84	265
bZIP	<u>b</u> asic leucine <u>z</u> ipper	127	218	140	66	266
WRKY	présence d'un motif <u>W</u> (tryptophane) <u>R</u> (Arginine) <u>K</u> (lysine) <u>Y</u> (Tyrosine)	90	163	128	81	233
B3	domaine de type B3	77	77	65	85	112
C3H	<u>C</u> ys-X8- <u>C</u> ys-X5- <u>C</u> ys-X3- <u>H</u> is	66	111	74	44	136
G2-like	domaine GARP dérivé de domaines de liaison à l'ADN chez plusieurs espèces: GOLDEN2 du maïs, ARR d'Arabidopsis, et Psr1 de Clamydomonas	64	89	62	35	164
GRAS	dérivé de <u>G</u> AI (gibberellin(GA) insensitive), <u>R</u> GA (repressor of GA) and <u>SCR</u> (SCARECROW)	37	104	69	61	139
HD-ZIP	<u>H</u> omeodomain leucine <u>z</u> ipper	58	97	61	34	140
M-type	MADS (MCM1-agamous-ARG80-deficiens-SRF) box	70	47	35	73	88
MIKC	dérivé de <u>M</u> ADS box, <u>i</u> ntervening domain, <u>k</u> eratin-like coiled-coil domain and <u>C</u> terminal domain	76	90	61	16	160
LBD	<u>L</u> OB (lateral organ boundaries) <u>b</u> inding <u>d</u> omain	50	60	39	33	111
Trihelix	helix-loop-helix-loop-helix	34	59	40	28	93
FAR1	dérivé de FAR-RED ELONGATED HYPOCOTYL3 (FHY3) et FAR-RED IMPAIRED RESPONSE1 (FAR1)	26	25	133	124	103
Dof	<u>D</u> NA binding with <u>o</u> ne <u>f</u> inger	47	51	37	26	93
GATA	reconnaissance du motif WGATAR (avec W=A ou T et R=A ou G)	41	54	32	36	70
ARF	<u>A</u> uxin <u>r</u> esponse <u>f</u> actor	37	62	48	26	85
HSF	<u>H</u> eat <u>s</u> tress transcription <u>f</u> actor	25	49	38	21	61
TALE	<u>t</u> hree- <u>a</u> mino-acid-loop- <u>e</u> xtension, classe d'homeoprotéine	33	52	45	13	101
AP2	2 répétitions du domaine AP2/ERF (APETALA2/ethylene response factor)	30	54	22	20	76
TCP	<u>t</u> eosinte branched1 (tb1-maize), <u>C</u> YCLOIDEA (CYC-snapdragon) and the <u>P</u> ROLIFERATING CELL FACTORS 1 and 2 (PCF1 and PCF2-rice), bHLH family	33	52	23	15	71

Tableau I-1 – Liste des 25 types de facteurs de transcription les plus courant d'après les données disponibles sur <http://planttfdb.cbi.pku.edu.cn/> (dernière mise à jour au 23-08-2013). Répartition dans différentes plantes modèles: *A. thaliana*, maïs, riz, *M. truncatula* et soja.

transcription. Cette tendance (*i.e.* activation/répression) est déterminée par l'intervention de différents partenaires (Brivanlou and Darnell, 2002). De nombreux acteurs peuvent ainsi influencer la régulation par les FTs et notamment les événements post-transcriptionnels comme les acétylations ou les phosphorylations/déphosphorylations. Ces dernières ont une grande influence sur l'activité des FTs car elles peuvent entraîner des modifications dans la liaison à l'ADN, engendrer des changements dans les signaux d'adressage vers le noyau ou encore altérer l'activité de régulation par les FTs (Schwechheimer and Bevan, 1998).

I.1.1 - Diversité fonctionnelle des familles de FTs les plus répandues

Les FTs se déclinent en plusieurs familles identifiées grâce à leur domaine de liaison à l'ADN. Une mise à jour récente de la base de données spécifique des FTs de plantes (<http://plantfdb.cbi.pku.edu.cn/>) porte leur nombre à 129 288 (Jin et al., 2014). En se basant sur ces informations, le tableau I-1 reprend les 25 classes de FTs les plus abondantes au sein des 83 espèces répertoriées sur le portail Plant Transcription Factor Database (Zhang et al., 2011; Jin et al., 2014), avec leur description, ainsi que le nombre identifié pour plusieurs espèces modèles présentées (*A. thaliana*, *Zea mays*, *Oryza sativa*, *M. truncatula* et *Glycine max*). On observe une importante similitude dans l'abondance des FTs au sein des espèces présentées, avec notamment les TFs de type bHLH et MYB les plus nombreux. Bien que l'on constate des disparités entre les valeurs énoncées dans le tableau et celles mentionnées dans l'article de Feller et al. concernant *A. thaliana* et le riz, ces deux familles de FTs restent les plus répandues et sont caractérisées par un large spectre de régulation (Feller et al., 2011).

Les FTs de type bHLH (basic helix loop helix) sont généralement classés en fonction des spécificités de leur domaine de liaison à l'ADN et de la présence ou absence de domaines additionnels conservés. Quatre groupes avaient ainsi été définis lors des premiers arbres phylogéniques. Ce nombre se porte aujourd'hui à 32 avec des variantes en fonction de l'identification de nouveaux domaines conservés (Feller et al., 2011). Ces FTs de type bHLH interviennent dans de multiples processus physiologiques comme la répression de la germination *via* le FT SPATULA qui permet de contrôler ce processus en réponse à des changements environnementaux (lumière et température) (Penfield et al., 2005). Le FT TT8 (TRANSSPARENT IESTA8) serait impliqué dans les voies de biosynthèse de flavonoïdes, conférant la pigmentation des graines (Nesi et al., 2000). Les FTs MYB se déclinent en trois groupes (MYB-R1, MYB-R2R3 et MYB-R1-R2-R3) en fonction du nombre de répétition de leur motif variable R, les membres de la famille MYB-R2R3 étant les plus répandus chez les plantes. Ces FTs MYB-R2-R3 ont été décrits comme impliqués dans la régulation de la synthèse de pro-anthocyanine ou d'anthocyanine chez *M. truncatula*: *MtPAR* (Verdier et al., 2012) ou chez le soja: Glyma09g36990 (Gillman et al., 2011) ou encore la réponse aux stress hormonaux chez le riz: OsMYBS1-3 (Lu et al., 2002). Ces deux familles de TFs bHLH et MYB agissent également en synergie avec la famille de protéines WD40, et ce en activant ou en réprimant des voies métaboliques liées au développement des

cellules de l'épiderme, comme la production d'anthocyanine, la formation des trichomes ou encore la production de mucilage (Ramsay and Glover, 2005; Feller et al., 2011).

Les FTs de type ERF (Ethylene Response Factor) appartenant à la famille des AP2 (Apetala-2)/ERF sont également fortement représentés parmi les FTs les plus fréquemment retrouvés chez les plantes ; ils sont dans la majorité des cas associés à des processus de réponse aux stress. Chez le riz, l'expression du FT TSRF1 (tomato stress responsive factor) semble contrôler de nombreux paramètres physiologiques au sein de la plante, et notamment l'amélioration de sa tolérance à la sécheresse. Son expression induit l'activation de gènes de réponse aux stress en augmentant la sensibilité des plantes à l'ABA (Acide ABscissique). Une accumulation de proline et de sucres solubles dans les grains est également observée chez les plantes sur-exprimant ce FT (Quan et al., 2010). Chez le maïs, la surexpression simultanée de ce même FT TSRF1 et d'une protéine de pomme de terre riche en lysine SBGLR (*Solanum tuberosum* genomic lysine-rich) entraîne une augmentation de la teneur en lysine et en protéine brute des graines, de même qu'une forte surexpression de gènes de réponse aux stress salins (Wang et al., 2013).

Les FTs de types NAC sont également impliqués dans les voies de réponse au stress environnementaux comme la sécheresse (Lu et al., 2012), mais sont aussi associés à des processus physiologiques programmés comme la sénescence des feuilles. En effet, l'étude du FT OsNAP chez le riz montre une forte expression lors de la sénescence des feuilles ou d'un traitement par le méthyl-jasmonate. Des lignées sur-exprimant ce FT présentent une sénescence accélérée dû à une augmentation de la biosynthèse de jasmonate (Zhou et al., 2013).

La famille de FTs de type WRKY rassemblent des gènes impliqués dans les voies de réponse à des agressions par des pathogènes ou induits par des signaux hormonaux. Ces FTS interviennent également dans la biosynthèse des métabolites secondaires et la formation des trichomes. Chez *A. thaliana*, WRKY23 intervient dans la distribution d'auxine dans les racines (régulation négative du transport) en stimulant la synthèse de flavonols (Grunewald et al., 2012). Ce gène semblerait également être impliqué dans le développement de la graine, car des mutants RNAi présentent des troubles dans la distribution d'auxine dans l'embryon ce qui conduit à des défauts de développement (Grunewald et al., 2013). Les interactions entre les TFs et les hormones au niveau des méristèmes apicaux et racinaires sont en effet déterminantes pour le devenir des cellules et leurs différenciations (Long and Benfey, 2006).

On remarque d'après le tableau que les facteurs de transcription de type FAR (Far-Red impaired Response) sont très présents chez le riz et *M. truncatula* alors que très peu de gènes FAR pour ces deux espèces ont été décrits dans la littérature. D'abord identifiée chez *A. thaliana* grâce à un mutant *far* présentant une réduction dans sa réactivité face à une exposition continue à des longueurs d'ondes rouge-lointaine, FAR1 a été décrite comme une protéine contenant un signal d'adressage au noyau. L'absence ou l'inactivité de FAR1 conduit à une perte de signaux au sein du phytochrome A (phyA), ce qui laisse supposer que cette protéine est essentielle dans la transduction du signal lumineux (Hudson et al., 1999). En

convertissant l'énergie lumineuse en énergie chimique, les phytochromes jouent un rôle important dans de nombreux processus physiologiques comme l'induction florale, la régulation du cycle circadien ou encore la biosynthèse des pigments. Un lien avec le métabolisme hormonal a également été suggéré. Ainsi FHY3 (Far-red elongated hypocotyl3) et FAR1, qui régulent positivement la transduction du signal dans le phytochrome A (Li et al., 2010) seraient également impliqués dans la perception de l'ABA pour l'induction de la germination (Tang et al., 2013).

Parmi les différentes familles de FTs présentées, on constate que les FTs sont impliqués dans de nombreux processus de développement de la plante (germination, réponse au stress, développement racinaire, réponse hormonale..). On retrouve également de nombreuses études portant sur la régulation du développement de la graine par les FTs (Le et al., 2007; Verdier et al., 2008; Wang et al., 2010; Agarwal et al., 2011).

I.1.2 - FTs impliqués dans le développement de la graine

L'analyse des FTs impliqués dans le développement de la graine pourrait permettre de mieux comprendre les mécanismes qui interviennent au cours des différents stades de développement et de mettre en avant des spécificités tissulaires. Quelques exemples de FTs impliqués dans le développement de la graine ont été classés selon leur spécificité tissulaire et les stades pendant lesquels ils montrent une expression préférentielle (partie 5.2. Transcriptional regulation (Zuber et al., 2012)). Ainsi, les études des FTs chez *A. thaliana* ont permis de mettre en évidence une variabilité importante dans la distribution des FTs au cours du développement de la graine et dans ses différents tissus (Le et al., 2007; Day et al., 2008). On note cependant que l'on retrouve peu les familles bHLH et MYB parmi ces FTs spécifiques de graines. Néanmoins, certains FTs isolés de la famille bHLH constituent des exemples probants de FTs impliqués dans le contrôle de la taille de la graine et/ou du fruit (taille de l'embryon : RGE1 Retarded growth of embryo 1 (Kondou et al., 2008), expansion cellulaire : VvCEB1 (Nicolas et al., 2013)).

Parmi la distribution des FTs au cours du développement de la graine chez *A. thaliana* (Le et al., 2007), on note une occurrence plus importante des FTs de type MADS au stade globulaire, suggérant leur implication dans la phase d'embryogenèse. Un FT de type MADS a récemment été identifié chez le riz comme déterminant dans la structure et la morphologie de l'albumen et l'embryon (Nayar et al., 2013). Dans une analyse similaire pendant le développement embryonnaire dans la graine de colza, les ESTs correspondant aux FTs de type MADS, YABBY, MYB et HDZIP ont été dénombrés et leur distribution a été étudiée sur plusieurs stades de développement du stade zygote au stade mature (Venglat et al., 2013). Cette étude a également montré que les ESTs correspondants aux MADS sont très fortement représentés uniquement au stade zygote, confortant les observations précédentes. Venglat et al. ont également montré que les ESTs codant des FTs de type MYB sont majoritairement présents pendant les stades très précoces du développement de l'embryon, avec plus de 30 ESTs dénombrés au stade où l'embryon est

composé de quatre cellules puis leur nombre diminue fortement, suggérant ainsi un rôle dès les premières divisions cellulaires. Certains FTS de cette famille ont en effet été décrits comme impliqués dans l'embryogenèse. Chez *A. thaliana*, le FT AtMYB118, exprimé préférentiellement dans l'embryon des stades précoces du développement jusqu'au stade maturation, pourrait réguler l'expression des protéines LEA ainsi que d'autres protéines de l'embryon et jouer ainsi un rôle important dans l'embryogenèse (Zhang et al., 2009). De même, le FT MYB56 exprimé dans les graines du stade cœur au stade cotylédons, semble également impliqué dans le contrôle de la taille de la graine en régulant l'expansion cellulaire (Zhang et al., 2013). Une étude des ESTs chez le maïs (Wang et al., 2010) qui regroupe de façon plus globale l'ensemble des phases de développement de la graine de 3 jap (stade qui regroupe la double fécondation, la formation du syncytium et la cellularisation) à 36 jap (stade de dessiccation intense) montre que les familles de FTs majoritairement retrouvées sont les Zn-finger et bZIP. Cette analyse indique également la présence significative de plusieurs autres types de FTs qui seraient impliqués dans ces stades de l'embryogenèse à la maturation, il s'agit des familles bHLH, MADS, MYB, Homeodomain et AP2/EREBP. Une analyse complémentaire des profils d'expression de ces FTs au cours des différents stades de développement par q-PCR a ensuite montré une prédominance des FTs de type MADS en tout début d'embryogenèse (déjà énoncé plus haut) alors que certains FTs Zn finger seraient préférentiellement exprimés pendant l'embryogenèse mais aussi lors du remplissage (Wang et al., 2010). Ces différentes études confirment donc une implication importante des FTs MADS plutôt en début de développement de l'embryon, alors que certaines familles de FTs seraient exprimées plutôt constitutivement ou ponctuellement au cours du développement.

1.2 - Focus sur les facteurs de transcription de type DOF (revue)

The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants.

Mélanie Noguero^a, Rana Muhammad Atif^a, Sergio Ochatt^a, Richard D. Thompson^{a*}

^aINRA, UMR1347 Agroécologie, BP 86510, F-21000 Dijon, France

*Corresponding author.

E-mail address: thompson@dijon.inra.fr

Plant Science 209 (2013) 32– 45

1.2.1 - Abstract

The DOF (DNA-binding One Zinc Finger) family of transcription factors is involved in many fundamental processes in higher plants, including responses to light and phytohormones as well as roles in seed maturation and germination. DOF transcription factor genes are restricted in their distribution to plants, where they are in many copies in both gymnosperms and angiosperms and also present in lower plants

such as the moss *Physcomitrella patens* and in the alga *Chlamydomonas reinhardtii* which possesses a single DOF gene. DOF transcription factors bind to their promoter targets at the consensus sequence AAAG. This binding depends upon the presence of the highly conserved DOF domain in the protein. Depending on the target gene, DOF factor binding may activate or repress transcription. DOF factors are expressed in most if not all tissues of higher plants, but frequently appear to be functionally redundant. Recent next-generation sequencing data provide a more comprehensive survey of the distribution of DOF sequence classes among plant species and within tissue types, and clues as to the evolution of functions assumed by this transcription factor family. DOFs do not appear to be implicated in the initial differentiation of the plant body plan into organs via the resolution of meristematic zones, in contrast to MADS-box and homeobox transcription factors, which are found in other non-plant eukaryotes, and this may reflect a more recent evolutionary origin.

Keywords: transcription factor, zinc finger, DOF, repressor, activator, stress, seed

I.2.2 - Discovery

Transcriptional regulation in higher plants is based on processes common to all multicellular organisms. The proteins implicated are generally recognizable on the basis of sequence homologies with orthologs in other eukaryotes. A number of plant-specific families of transcription factors exist, including the DOF (DNA-binding One Zinc Finger) family. The first gene containing a DOF domain was isolated from maize by screening a lambda gt11 expression library of maize leaf cDNA with a radiolabelled promoter fragment from the CAMV 35S promoter (Yanagisawa and Izui, 1993). DOF transcription factors were rapidly identified in several species, playing a variety of roles (Kisu et al., 1995; Zhang et al., 1995; DePaolis et al., 1996; Vicente-Carbajosa et al., 1997). DOF genes occur in all angiosperms, gymnosperms, and lower plants including the green alga *Chlamydomonas reinhardtii* (*C. reinhardtii*) and the moss *Physcomitrella patens* (*P. patens*). DOF transcription factors are implicated in a wide range of processes, having roles such as regulators of stress responses, photosynthesis or flower induction. They are clearly vital for plant development in a number of cases. DOFs have been reported to activate or to repress transcription, depending on the factor concerned and the gene target. The advent of next generation sequencing has vastly increased the number of DOF gene sequences available, and gives access to a larger dataset of the associated expression profiles, rendering an updated review of the distribution and roles of this transcription factor family opportune.

I.2.3 - Phylogeny of DOF genes

Different strategies have been used to discern the phylogenetic relationships between members of the DOF gene family. A phylogenetic tree was constructed based on 30 rice *Oryza sativa* (*O. sativa*) and 36 *Arabidopsis thaliana* (*A. thaliana*) DOF sequences taken from the genomic sequences of the two organisms

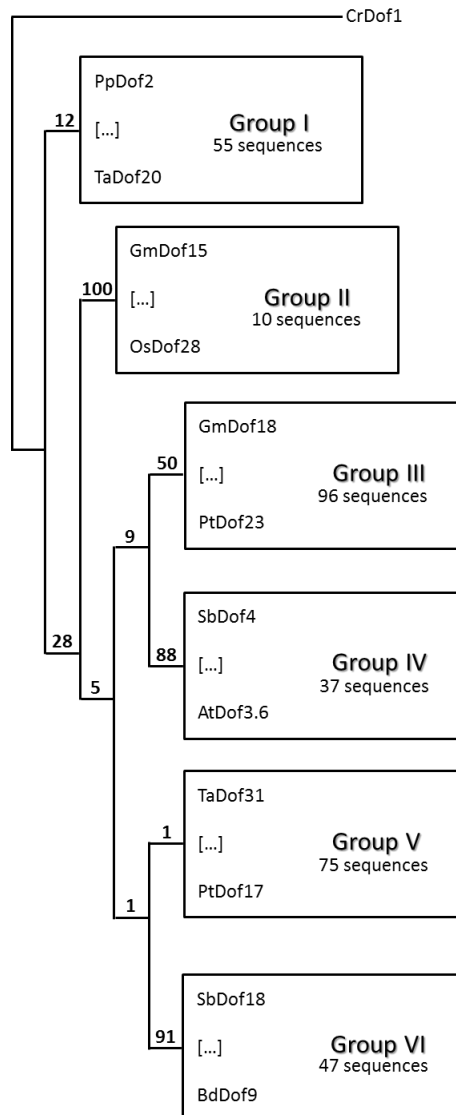


Figure I - 1: The phylogenetic tree was created using the full length DOF protein sequences from *A. thaliana* (36 AtDof) (Yanagisawa, 2002), *B. distachyon* (27 BdDof) (Hernando-Amado et al., 2012), *C. reinhardtii* (1 CrDof) (Moreno-Risueno et al., 2007), *G. max* (28 GmDof), *H. vulgare* (24 HvDof) (Moreno-Risueno et al., 2007), *M. truncatula* (22 MedtrXg), *O. sativa* (31 OsDof) (Kushwaha et al., 2011), *P. patens* (19 PpDof) (Shigyo et al., 2007; Sugiyama et al., 2012), *P. trichocarpa* (61 PtDof) (<http://www.plantgdb.org/>), *P. sativum* (7 PsDof), *S. bicolor* (28 SbDof) (Kushwaha et al., 2011), *S. moellendorffii* (5 SmDof) (Moreno-Risueno et al., 2007), *T.aestivum* (23 TaDof) (Shaw et al., 2009) and from *Z.mays* (10 ZmDof) (<http://www.maizegdb.org>). These 322 DOF protein sequences (detail in supplemental table 1) were aligned and organized in a tree based on sequence similarities (gap open cost: 10 and gap extension cost: 1) with CLC sequence viewer 6.2, selecting neighbor-joining as the phylogenetic reconstruction method. The tree was rooted using Cr-Dof1 as an out-group. Six groups were identified based on sequence similarities. The composition of each of these groups is presented in supplemental figures 1 to 6.

and therefore also including intron sequences. The alignment of the deduced protein sequences was done using ClustalW, and was however based on the DOF domain alone. A joined, unrooted phylogenetic tree of the two datasets, inferred by nearest neighbour alignment (Lijavetzky et al., 2003) showed the presence of four main clusters of orthologous genes (MCOG). The presence of conserved motifs among the main clusters of orthologous genes was investigated using MEME (Multiple Expectation-Maximization for Motif Elicitation) program (Bailey et al., 2009), (<http://meme.nbcr.net/meme/cgi-bin/meme.cgi>). Most of the motifs found have no known function, but one contained a predicted nuclear localization signal. By adding DOF gene sequences from lower plants (1 from *C. reinhardtii*, used to root the tree, 9 from *P. patens*, 8 from the fern *Selaginella moellendorffii* (*S. moellendorffii*), 8 from the gymnosperm *Pinus taeda* (*P. taeda*), and 26 from barley *Hordeum vulgare* (*H. vulgare*)), the analysis was refined (Moreno-Risueno et al., 2007). Using 116 sequences, a phylogenetic tree was constructed using Bayesian inference (Bayes 3.1) which revealed 7 sub-families or MCOGs supported by bootstrap values of >50. Most of the DOFs from lower plants group close to the root of the tree, within one subfamily. Once again, MEME was used to predict protein motifs characteristic of the different DOF subgroups. Most of these motifs were restricted to one or a few subgroups. Two motifs corresponding to possible nuclear localization signals were identified, located just N-terminal of the DOF domain.

For the present review, a larger dataset (>300 sequences) has been analyzed (Fig. I-1). We have taken the available DOF protein sequences for *A. thaliana* (36 AtDof), *Brachypodium distachyon* (27 BdDof), *Glycine max* (28 GmDof), *H. vulgare* (24 HvDof), *Medicago truncatula* (22 MedtrXg), *O. sativa* (31 OsDof), *Populus trichocarpa* (61 PtDof), *Pisum sativum* (7 PsDof), *Sorghum bicolor* (28 SbDof), *Triticum aestivum* (23 TaDof) and *Zea mays* (10 ZmDof) (see supplemental Table 1 for sequences).

We have also included possible ancestral genes from the green unicellular alga *C. reinhardtii*, the fern ally *S. moellendorffii* and the moss *P. patens* (five for *S. moellendorffii* (SmDof) (Moreno-Risueno et al., 2007), nineteen sequences for *P. patens* (PpDof) (Shigyo et al., 2007; Sugiyama et al., 2012) and a single DOF gene (CrDOF1), found for *C. reinhardtii* (Moreno-Risueno et al., 2007)).

These 322 DOF protein sequences (detail in supplemental table 1) were aligned and organized in a tree based on sequence similarities (gap open cost: 10 and gap extension cost: 1) with CLC sequence viewer 6.2. We used the accurate built-in algorithm of CLC sequence viewer and selected neighbor joining as the phylogenetic reconstruction method (Gaur and Li, 2000). The tree structure, rooted using Cr-Dof1 as an outgroup is presented in Fig. I-1. Six groups were identified based on sequence similarities despite low bootstrap values. The composition of each group is presented in supplemental figures 1 to 6.

The robustness of the phylogenetic tree is limited by the fact that DOF sequences consist of a very highly conserved DOF domain embedded in the remaining sequence showing little evidence of selection. Six groups were identified based on sequence similarities but due to low bootstrap values for some of these groups they should be considered with reserve.

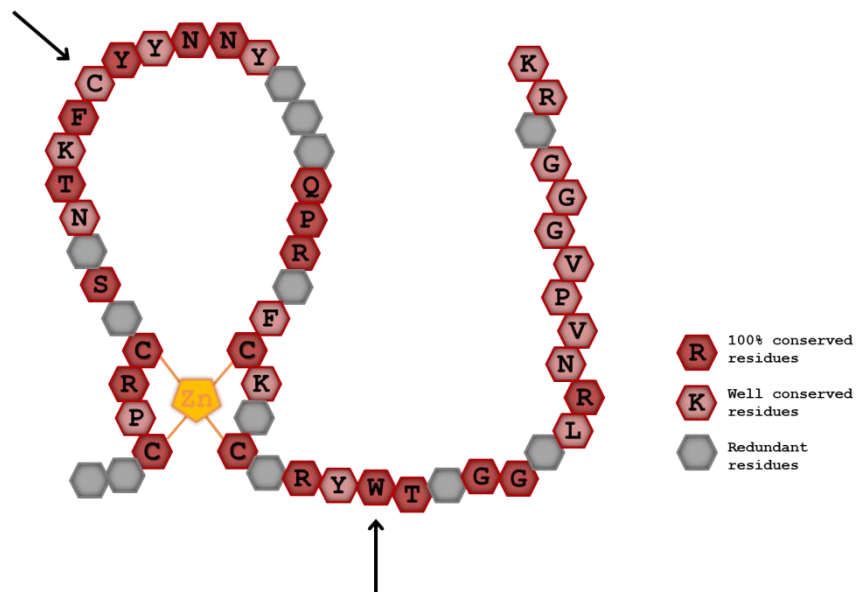


Figure I-2: Schematic representation of the structure of the DOF domain. Conservation was deduced by aligning the 52 amino acid domain from all available DOF protein sequences from rice, barley, green unicellular alga, moss, fern, sorghum, brachypodium, pea, soybean, *A. thaliana*, *M. truncatula*, wheat, maize and poplar (all sequences used for Figure 2). Absolutely conserved amino acid residues are shown in red. Well conserved residues (up to 75 % of conservation) are shown in pink and less-conserved residues are shown in grey. Two amino acid residues discussed in section 4 of the text (the non-Zn finger cysteine and the tryptophan at the C-terminal of the DOF domain) are indicated by arrows.

The lower plant DOF sequences which were added to the phylogenetic analysis to deduce possible evolutionary relationships are present in three groups (groups I, III and V) (see supplemental Fig 1 to 6). While some of the sequences form discrete sub-groups (cluster of 6 PpDof and 4 SmDof sequences in group I showed by square brackets in supplemental Fig.1, and cluster of 11 PpDof sequences in group V highlight by square bracket in supplemental Fig.5), others are more similar to DOF sequences from higher plants: PpDof18 and SmDof5, identified by arrows in supplemental Fig.3 (relatively close each other with a bootstrap value of 86) in group III and PpDof19, marked by arrow in supplemental Fig. 5 (closer to GmDof26 with a bootstrap value of 67) in group V. Overall, the tree supports a monophyletic origin despite low bootstrap values for certain branches, and is broadly consistent with earlier analyses. A phylogenetic study of the DOF domain with sequences from *A. thaliana*, rice and *P. patens* found a group A similar to group I; a group B which is a subgroup in group V, and part of the group C is in group IV (Shigyo et al., 2007). Another study of *A. thaliana*, rice and sorghum sequences in which the complete DOF protein sequences were analyzed also shows some clusters in common with our analysis. Group A is similar to group I and group C is shared between group V and VI; whilst group E is close to group IV (Kushwaha et al., 2011). A further analysis of rice and *A. thaliana* DOF genes which included DOFs from lower plants to root the tree generated a tree which was globally similar to that presented in this study (Moreno-Risueno et al., 2007).

I.2.4 - Definition of the DOF domain and its DNA target sites

DOF transcription factors contain the DOF DNA-binding domain usually located close to the N-terminal region of the protein. The highly conserved DOF domain is a region of 52 amino acid residues containing the string CX₂CX₂₁CX₂C, which binds zinc (Zn²⁺) in a characteristic zinc finger configuration (Fig. 2). This motif is essential for DNA binding and mutation of any of the four cysteines produces dramatic changes in the finger configuration, and abolishes DNA binding in vitro and in vivo (Yanagisawa, 2001). Both the loop length and the presence of a Zn²⁺ ion directly affect the loop structure and determine the nature of the DNA binding. A study of the pumpkin DOF AOBP, which binds to the gene encoding ascorbate oxidase, showed that the four cysteine positions that constitute the zinc finger are also required for loop stability. The tryptophan at the C-terminal base of the zinc finger is also important for DNA binding (Shimofurutani et al., 1998), and in steroid hormone receptors this residue appears to stabilize the structure of the domain (Kisu et al., 1998).

The core recognition site of DOF proteins is AAAG (or its reversibly complementary sequence, CTTT), except in pumpkin, where a variant binding site, AAGT, was found (Kisu et al., 1998). The sequences flanking the AAAG binding site may also influence DOF binding to DNA, and not all AAAG strings are DOF binding sites (Yanagisawa and Schmidt, 1999). DOF proteins can bind to promoters containing either a

unique AAAG motif or more frequently, repetitions of this motif (Yanagisawa, 2000). The binding of DOF proteins to the AAAG motif has been confirmed for selected promoter sequences by electrophoretic mobility shift assays in vitro (Yanagisawa, 2002; Umemura et al., 2004). The regions N- and C-terminal to the DOF domain, unlike the DOF domain itself, are very variable among different DOF proteins. These regions may interact with different regulatory proteins or intercept signals that mediate activation or repression of gene expression. These interactions are likely to contribute to the diverse functions of DOF domain proteins (Yanagisawa, 2001). In addition to the DOF DNA-binding domain, a sequence necessary for nuclear localization has been identified in certain DOF proteins, and a bipartite nuclear localization signal, which overlaps partly with the conserved DOF domain, has been mapped (Yanagisawa, 2001; Krebs et al., 2010). Furthermore, a transcriptional activation domain was identified in the C-terminus of the *Zea mays Dof1* gene (Yanagisawa, 2001). In addition, DOF protein domains, including the DOF domain itself, are also involved in physical interactions between DOFs and a number of other transcription factors (16, 17, see next section).

1.2.5 - Physical and functional association of DOFs with other transcription factors

As the canonical recognition sequence for DOF transcription factors is very short, it is highly likely that they function in association with other transcription factors and the combination generates the necessary promoter specificity. There are several reports of interactions between DOF proteins and other transcription factors. A DOF protein (OBP1) interacts with bZIP transcription factors OBF4 and OBF5 (Zhang et al., 1995). The DOF OBP1 facilitates the binding of the associated bZIP OBF transcription factor to its DNA target, the OCS element, although not binding stably by itself (Zhang et al., 1995). AtDOF4.7 interacts with a member of another Zn-finger TF family, AtZFP2, together implicated in the control of floral organ abscission (Wei et al., 2010). Maize Dof1 protein interacts with another DOF, Dof2, and with High Mobility Group (HMG) proteins in the absence of DNA, and also in the presence of a DOF binding site oligonucleotide (Yanagisawa, 1997). The HMG proteins may be acting as chaperones that assist DOF factor binding to DNA (Krohn et al., 2002). These protein-protein interactions require the presence of the DOF domain, but not DNA-binding, as a mutation of one of the cysteines required for Zn-finger formation did not abolish the Dof1-Dof2 or the Dof1-HMG protein interactions.

The promoters of genes expressed during cereal endosperm development share a conserved motif, the prolamin or endosperm box (P-box), which is sufficient to confer endosperm-specific expression on a minimal promoter. This motif is bipartite, with part being recognized by a genetically characterized bZIP transcription factor, the product of the *Opaque-2* locus in maize. The remaining AAAG string of the P-box is the site of binding of a maize DOF transcription factor, termed prolamin box binding factor (PBF). Both *Opaque-2* and PBF are specifically expressed during maize endosperm development (Vicente-Carbajosa et al., 1997) and the proteins bind to each other in vitro. The putative DOF PBF ortholog in barley, BPBF, also.

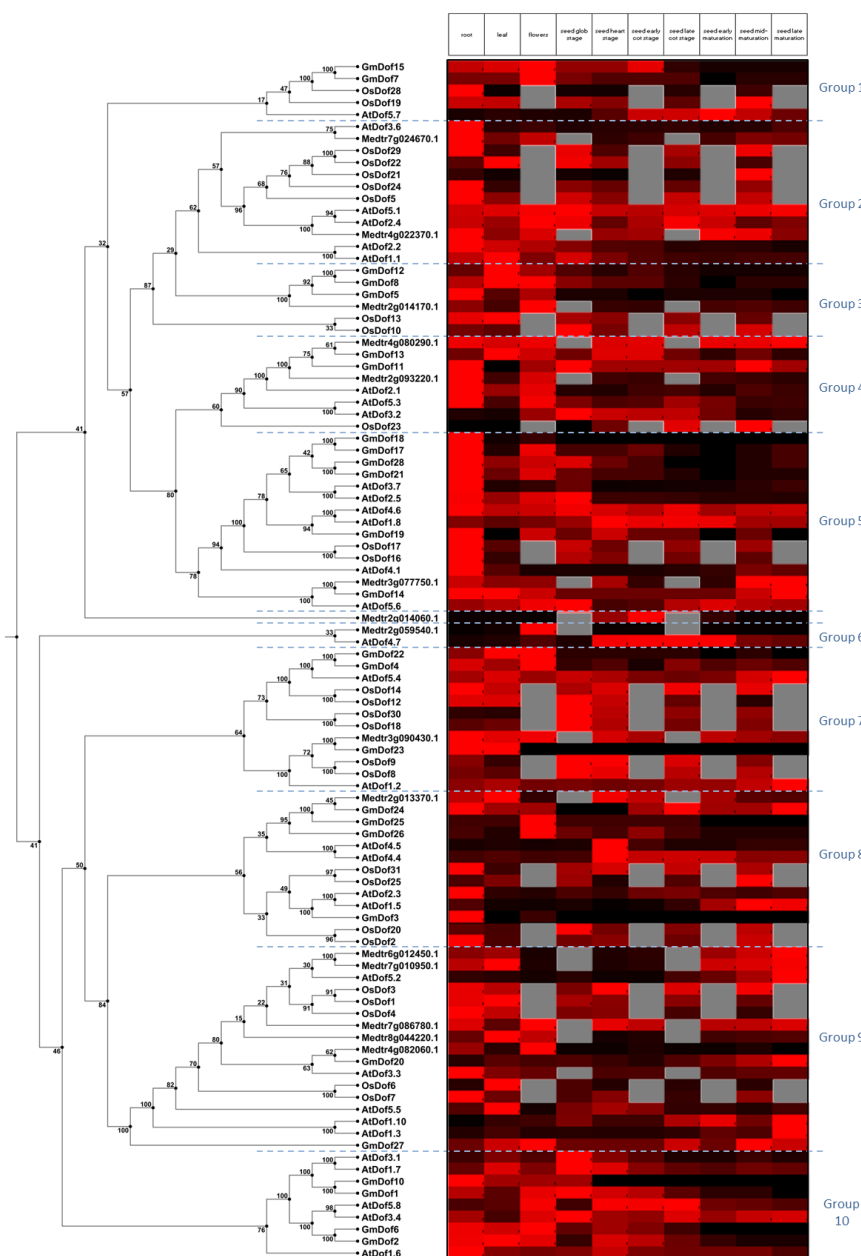


Figure I-3: A simplified phylogenetic tree was created using the full length DOF protein sequences: 32 from *A. thaliana* (At), 26 from *G. max* (Gm), 15 from *M. truncatula* (Medtr) and 26 from *O. sativa* (Os) for which expression profiles have been published. These 99 DOF domain protein sequences were aligned and organized in a tree based on sequence similarities (gap open cost: 10 and gap extension cost: 1) with CLC sequence viewer 6.2. Expression profiles were derived from TAIR AtGenExpress Visualization Tool, Soybase.org, eFP browser (<http://bar.utoronto.ca>) and mtgea.noble.org for soybean, *A. thaliana*, rice and *M. truncatula*, respectively. The data were taken from <http://www.weigelworld.org/> for *A. thaliana*, <http://mtgea.noble.org/v3/index.php> for *M. truncatula*, Jain et al. 2007 for rice (Jain et al., 2007), and <http://soybase.org/> for soybean (see supplemental Table 1 for tissue organization). All data were re-normalized from 0 to 1 (for the highest value) for each gene (supplementary table 3), species by species, to conserve the differential expression between tissues, and were then compiled and reorganized to match the order generated by the phylogenetic tree.

interacts with a Myb transcription factor, GAMyb. In this case, both the DOF and the Myb DNA binding sites are required on the target promoter for transcriptional activation (Diaz et al., 2002)

I.2.6 - Relationship of DOF proteins to GATA and steroid hormone receptors

Although no structural information is currently available for the DOF protein domain, the domain has some sequence similarities to two other families of C2C2 Zn finger transcription factors, the GATA family and the steroid hormone receptor family (Umemura et al., 2004), which have provoked speculation about functional homologies (Moreno-Risueno et al., 2007; Shigyo et al., 2007). Besides the presence of four similarly spaced cysteine residues (CXXC (X)_n...CXXC), separated by a loop of 21 amino acids in the case of DOFs, two adjacent aromatic residues are completely conserved just downstream of the second Zn finger Cys pair (YW in DOF proteins). Substitution of the tryptophan residue (identified with an arrow on Fig.2) of the DOF pumpkin ascorbate oxidase binding protein (AOBP) by non-aromatic residues reduces (in the case of histidine) or eliminates (for alanine) DNA-binding depending on the replacement residues used (Umemura et al., 2004). The greater loop length of the DOF domain in comparison to the DNA-binding site of the GATA factors or steroid hormone receptor domains seems important for DNA binding, as shortening the loop by six residues, or replacement of the loop by the loop region from zinc binding region 1 of the estrogen receptor both eliminated specific DNA binding. The specificity of DNA binding is thought to be conferred by the alpha-helical region at the C-terminal end of the DOF domain, but the loop is needed to maintain a configuration capable of DNA binding (Shimofurutani et al., 1998). The presence of a fifth cysteine residue (highlighted with an arrow on Fig. I-2) within the DNA-binding loop of the DOF domain further distinguishes DOFs from GATA or steroid hormone transcription factors. The deletion of this cysteine increased binding affinity (Shimofurutani et al., 1998). Thus it may possibly be regulating DNA binding, which we speculate could provide a possible mechanism for redox regulation of DOF transcription factor activity. Interestingly, redox-directed regulation, possibly reflecting a response to the oxidative burst, is implicated in regulating the activity of other transcription factors, including the bZIP factors TGA9 and TGA10 in *A. thaliana* (Murmu et al., 2010), which are closely related to TGA4, a further redox-regulated bZIP factor which interacts with a DOF (Zhang et al., 1995; Despres et al., 2003).

I.2.7 - Classification of DOF gene sequences according to their expression profiles.

The gene expression profiles in different plant tissues of the DOF genes selected for phylogenetic analysis were compared in order to get an overview of the distribution of expression of DOF genes in higher plants. This analysis was carried out on representatives of major plant families for which expression data were available for several tissues. From the sequences of these species, a total of 99 DOF probeset IDs were recovered: 32 of the 36 *A. thaliana* genes, 15/22 for *Medicago truncatula* (*M. truncatula*), 26/28 for soybean and 26/31 for rice. The 99 sequences were aligned and re-organized in a tree based on sequence

CHAPITRE -1-

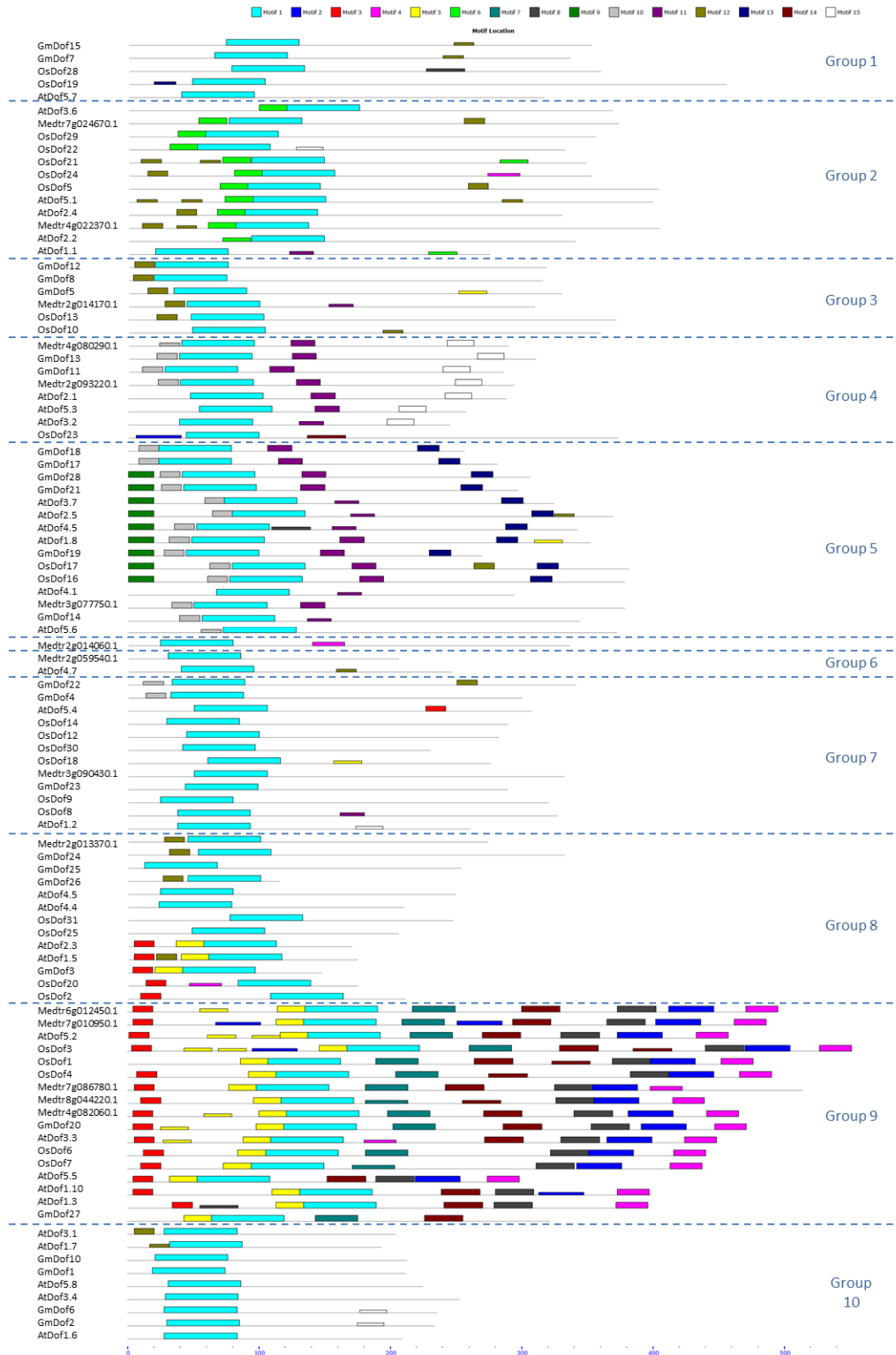


Figure I-4: Schematic graphical representation of distribution of conserved domains among the DOF proteins belonging to defined groups (from 1 to 10) as shown in Figure I-3. The domains were identified by means of the MEME software [7] using the complete amino acid sequences of the 32 *A. thaliana*, 26 *G. max*, 15 *M. truncatula* and 26 *O. sativa* DOF proteins. Positions of the identified motifs are shown according to scale. Consensus sequences of the MEME-defined motifs are listed in Table I-2.

similarities (gap open cost: 10 and gap extension cost: 1) with CLC sequence viewer, (Fig. I-3), using neighbor-joining as the phylogenetic reconstruction method (Gaur et al., 2011).

The expression data used were derived from public databases as indicated in the legend to Fig. I-3. Four tissues were examined: root, leaf, flowers and seed, for which different development stages were added, depending on the availability of the data (see supplemental Table 1 for tissue organization). All data were re-normalized from 0 to 1 (1 for the highest value) for each gene (supplementary Table 3), species by species, to conserve the differential expression between tissues, and were then compiled and reorganized to match the order generated by the phylogenetic tree.

The analysis (Fig. I-3) showed that some DOF sequence groups are predominantly expressed in one tissue, for example, group 1 in flowers, group 4 and 5 in roots, group 7 in flowers and early seed development, and group 9 in maturing seeds. However, for other groups such as group 8 or group 10, most sequence members are expressed in several tissues, and different sequences within a group are in some cases expressed in different tissues. For these groups, there was no clear relatedness between DOF sequence similarity and site of expression for a given sequence. It appears that in many cases, DOF transcription factors have been selected for regulating genes in a given tissue from a pool of already diverged sequences, and that the selection of a given DOF sequence for a given purpose has often differed between plant families.

I.2.8 - Identification of conserved domains in DOF protein sequences

The protein sequences used for the expression analysis (32 *A. thaliana*, 15 *M. truncatula*, 26 soybean and 26 for rice) were analyzed by MEME (Bailey et al., 2009), to detect possible common conserved domains in addition to the DOF domain (Lijavetzky et al., 2003; Moreno-Risueno et al., 2007; Kushwaha et al., 2011). The first analysis of this type (Yanagisawa, 2002) used ClustalW on a set of 36 *A. thaliana* DOF genes plus 15 other DOF sequences from a wide range of species. This analysis identified a series of motifs present most of which were confined to one phylogenetic group of DOFs. Three of these motifs are related in sequence and could represent nuclear localization signals, although being located just upstream of the DOF domain, they do not correspond to the NLS identified experimentally just downstream of the DOF domain (Krebs et al., 2010).

The new analysis reported here identified 15 conserved motifs: a graphical view of the motifs identified is presented in Figure I-4 and Table I-2 describes in more detail the motif analysis. This analysis reveals information about motif distribution between the groups: several motifs were restricted to one DOF sequence group, such as motif 6, only found in group 2, motifs 9 and 13 in group 5, and motifs 2, 4, 7, 8 and 14 in group 9. A search for protein domains of known function was performed using Prosite (Sigrist et al., 2002; de Castro et al., 2006; Sigrist et al., 2012) including patterns with a high probability of occurrence, (Table I-2). The conclusions that can be drawn are limited, as most of the motifs identified by MEME lack a
















		Multilevel consensus sequence of conserved domains	Length	Occurrence among the 100 sequences of the tree	Motif localization	Motif identification via Prosite	E-value
	Motif 1	L[KPN]CPRC[DN]STNTKFCYNNY[NS]L[SN]QPR[HY]FCK[AT]CRRYWT[KAG][TAS]LRNVPGG[C[CRS]R[KR]NK[RK]	55	100		Zinc finger Dof-type signature [CprCd-stntkfCyynnynsqprhfCkaCrRyWtkGG] N-glycosylation site [NLSQ] N-myristoylation site [GTIrNV] or [GCrkNK] Protein kinase C phosphorylation site [TIR]	1.5E-5089
	Motif 2	[ED]DK[ET][EN][GK][SC][VL][WL]VPKT[L]RIDDP[DE]E[AV]AKSSIW[TAS]TLGIK	34	14 (group 9)	Just after motif 8	Protein kinase C phosphorylation site [TIR]	3.6E-219
	Motif 3	KDP[AG]IKLFGKTIP[FL]P	15	21 (mainly in groups 8 and 9)	N-terminal of protein sequence	ATP/GTP-binding site motif A (P-loop) [AiklFGKT]	2.30E-102
	Motif 4	[VP][EG][TS]SP[VA][LM]QANPAA[LM]SRSQ[NST]F[HQ]E[TS][ST]	24	15 (group 9)	C-terminal of protein sequence	No hit	1.00E-98
	Motif 5	[EN][ET][ET][TQ]ET[DT][EST][SN][SE][QK][ED][KT][TV]LKKPDKI	21	20 (mainly in groups 8 and 9)	N-terminal of DOF domain sequence	Casein kinase II phosphorylation site [TetD] or [SsqE] Protein kinase C phosphorylation site [TIK]	2.60E-89
	Motif 6	[AI]R[PV][GVN]SM[ASTV][ED]RAR[LM]A[KNR][IV]P[LQ]PE[PAT][AG]	21	10 (group 2)	N-terminal of DOF domain sequence	Casein kinase II phosphorylation site [SmaE]	1.30E-56
	Motif 7	HPSL[KN]CN[AEG]TVL[TK]FG[SP]D[ST]PLCESM[AE]SVLN[LJ][AG][ED][KQ]	32	11 (group 9)		Protein kinase C phosphorylation site [SIK] N-glycosylation site [NATV]	5.50E-64
	Motif 8	PSS[SP]PSSSx[SC]SS[PS][NS]SPTLGKHSR[DE][EG]Sxx[EK]	29	16 (group 9)	Just before motif 2	Casein kinase II phosphorylation site [SrdE] or [SxxE]	1.70E-54
	Motif 9	MD[TA][AT][QKH]W[PAHT]Q[GE][FL][GVQ][VEL][VKM][KP][NPQ][EMPV][EK][EIP][MV]	19	9 (group 5)	N-terminal of protein sequence	N-myristoylation site [GVvkNE]	3.80E-45
	Motif 10	C[PS][AS]P[MS]xER[KR][AP]RPQ[KE][DE]	15	17 (mainly in group 5)	N-terminal of DOF domain sequence	No hit	8.20E-37
	Motif 11	[HY][GDQ][SG][QN]DL[NT]LA[FL][PA]R[LA][QE][KDG][QY][HS][LH]	18	15 (mainly in groups 4 and 5)		No hit	2.20E-31
	Motif 12	xM[EP][KP][PQ]xQxQQQQQQ	15	17		No hit	1.70E-30
	Motif 13	GRLFPF[EG]DLKQL[VS]SS	16	11 (group 5)		No hit	1.90E-29
	Motif 14	[QEN][EQ][QN]VEN[NS][CN][HN][GN][FP][PL][PN][GQ][LV][PQ][CY][FY]P[GS][PAV]PWP[VL][PT]W[NS]P	29	13 (group 9)		No hit	7.30E-31
	Motif 15	[EP][SN][RK]V[LF][WG]GFPWQ[ML]NG[DG][ST]N[GI][GM]E	20	7 (group 4)		No hit	1.50E-29

Table I-2: Multilevel consensus sequences of the MEME-defined domains observed among different DOF proteins from *A. thaliana*, *G. max*, *M. truncatula* and *O. sativa*. The search for identification of 15 motifs was performed on consensus sequence obtained as MEME results using Prosite (<http://prosite.expasy.org/>) with default parameters including patterns with a high probability of occurrence, the maximum number of occurrences of any motif within the complete dataset was set to 100, minimum and maximum width of motifs was set to 15 and 55 amino acid residues respectively.

Specie	Gene	NCBI or Gene bank reference sequence	Role/site of expression	Target gene if known	Responsive to	Repressor/activator/coactivator ^a	Reference
Tissue differentiation <i>Arabidopsis thaliana</i>	AtDof2.4; At2g37590	NM.129315.3	Vascular tissue development, procambial cells of leaf primordia, roots and embryos			A	Konishi and Yanagisawa [47]
	AtDof4.1; At4g00940 AtDof4.7; At4g38000	NM.116321.4 NM.119961.4	Organ size regulation Floral organ abscission	PGAZAT, polygalacturonase gene Revoluta		R R (interacts with AtZFP2)	Ahmad et al. [37] Wei et al. [21]
	AtDof5.1; AT5G02460	NM.120324.3	Establishment of leaf blade polarity			A	Kim et al. [38]
	AtDof5.8; At5g66940	NM.126093.1	Vascular tissue development, leaf primordia, developing embryos, and developing flower buds			A	Konishi and Yanagisawa [47]
	AtOBP1; AtDof3.4; At3g50410	NM.114901.3	Cell cycle regulation	CYCD3;3 (D-type cyclin) core cell cycle gene		A	Skirycz et al. [36]
	AtOBP3; AtDof3.6; At3g55370	NM.001203166.1	Growth regulator	ORG1, ORG2, ORG3 (OBP responsive gene) (bHLH TFs)	Salicylic acid-induced, auxin	A	Kang and Singh [56] and Kang et al. [58]
	CDF1 (cycling dof factor-1); AtDof5.5; AT5G62430.1	NM.125637.3	Repression of Flower induction	Constans		R	Song et al. [39]
	DAG1; AtDof3.7; AT3G61850	NM.116050.3	Hypocotyl elongation		phytochrome B/PIL5 signaling pathway	R	Gabriele et al. [66]
	HCA2 (high cambial activity); AtDof5.6; At5g62940.1	NM.125688	Vascular tissue development			A	Guo et al. 2009
	SCAP1; AtDof5.7; At5g65590	NM.125957.2	Stomatal maturation and functioning	K ⁺ channel protein, MYB60 transcription factor, pectin methylesterase		A	Negi et al. [53]
<i>Jatropha curcas</i>	JcDOF3	GQ256648	Flowering time		Light-mediated circadian-clock	R or A depending on circadian-photoperiod signals	Yang et al. [44]
<i>Nicotiana tabacum</i>	NtBBF1 (roiB domain B Factor)	AJ009594.1	Vascular system	RoIB	Auxin	A	Paolis et al. [54] and Baumann et al. [55]
<i>Solanum tuberosum</i>	StDof1	AJ242853.1	Guard cell-specific expression	KST1 (K ⁺ influx channel of guard cells)		A	Plesch et al. [52]
<i>Zea mays</i>	ZmDof1; MNB1A	NM.001112239.1	Anther development	Zm401, pollen-specific		R	Yanagisawa [19]; Chen et al. [34]

Table I-3: Examples of some processes which are known to involve regulation by DOF transcription factors, presented in the schematic of Figure II-5).

CHAPITRE -1-

Specie	Gene	NCBI or Gene bank reference sequence	Role/site of expression	Target gene if known	Responsive to	Repressor/activator/coactivator ^a	Reference
Seed development							
<i>Arabidopsis thaliana</i>	DAG1; AtDof3.7; AT3G61850	NM.116050.3	Seed germination			R	Papi et al. [64]; Gabriele et al. [66]
	DAG1; AtDof3.7; AT3G61850	NM.116050.3	Dormancy and seed germination		Chloroplast development	R (indirectly affect ELIP gene expression)	Rizza et al. [67]
	DAG1; AtDof3.7; AT3G61850	NM.116050.3		GA biosynthetic gene AtGA3ox1	Dark and phyB/PIL5 signaling pathway	A	Papi et al. [45]; Gabriele et al. [66]
	DAG2; AtDof2.5; AT2G46590	NM.130224.7			Light	R	
<i>Glycine max</i>	AtDof6; AtDof3.2; At3g45610	NM.114430.2	Seed germination			A	Gualberti et al. [65]
	GmDof4	DQ857254.1	Storage protein synthesis	Storage protein <i>CRA1</i>		R (interacts with an activator, TCP14)	Rueda-Romero et al. [72]
	GmDof11	DQ857261.1	Seed lipid content	Acetyl CoA carboxylase		R	Wang et al. [33]
<i>Hordeum vulgare</i>			Seed lipid content	Long-chain-acyl CoA synthetase		A	
	BPBF	AJ000991.1	Seed Germination	Cathepsin		A	
<i>Oryza sativa</i>	SAD (scutellum and aleurone-expressed DOF)	AJ312297.1	Seed Germination	Cathepsin		R	Zou et al. [69]
	OsDof3 (OsDof10); Os02g15350.1	AB028131.2	Seed germination	CPD3 type3-carboxypeptidase	GA3	A	Isabel-La Moneda et al. [68]
<i>Zea mays</i>	ZmPBF and paralogs	U82230.1	Endosperm development, storage protein synthesis	22 kDa and Gamma zeins, storage protein genes		C (interacts with GAMyb)	Washio [70] and Washio [71]
Regulation of metabolism							
<i>Arabidopsis thaliana</i>	AtOBP1.2.3; AtDof3.4; At3g50410	NM.114901.3			Auxin-induced	C (with a bZIP TF, Opaque-2)	Vicente-Carbajosa et al. [4]; Marzabal et al. [59]
	AtOBP2; AtDof1.1; At1g07640	NM.001035911.1	Regulates indole glucosinolate synthesis, phloem-associated	<i>CYP83B1</i> (?)	Herbivore and Me-jasmonate-induced	R	Kang et al. [56]
	AtDof4.2; At4g21030	NM.118221.1	Regulator of phenylpropanoid metabolism		Alternaria and Me-jasmonate-induced	A	Skirycz et al. [85]
	COG1 (cogwheel); AtDof1.5; AT1G29160.1	NM.102657	Phytochrome signaling (phyA- and phyB-signaling pathways)		Light	R or A depending on target gene	Skirycz et al. [84]
<i>Oryza sativa</i>	OsDof25; MNB1A; Os08g38220		Carbon and nitrogen metabolism		Nitrate transporter/Ammonium transporter	R	Park et al. [43]
<i>Physcomitrella patens</i>	PpDof1	BAL46025.1	Vegetative growth filaments		Environmental nutrient (carbon and nitrogen) conditions	R or A depending on target gene	Santos et al. [81]
<i>Zea mays</i>	ZmDof1; MNB1A	NM.001112239.1	Light-dependent modulation of carbohydrate metabolism	PEPC (phosphoenolpyruvate carboxylase), cyPPDK1 (C4-type pyruvate orthophosphate dikinase)	Light	A	Sugiyama et al. [10]
	ZmDof2	X79934.1	Non-leaf carbohydrate metabolism	<i>cypdk</i> and <i>pepcZm2A</i>	Expressed in non-green tissues	R	Yanagisawa and Sheen [76] and Yanagisawa [17]

Table I-3 (continued): Examples of some processes which are known to involve regulation by DOF transcription factors, presented in the schematic of Figure II-5).

known function, and those assigned a function are Prosite motifs of low sequence specificity which may occur by chance. These results do not support the hypothesis (Yang et al., 2006) that sequences surrounding the DOF domain have been recruited from unrelated protein sequences. We have also carried out BLAST searches using the DOF domain-surrounding sequences (unpublished) and find these invariably recognize only DOF gene sequences.

A comparison of the sites of expression of DOF sequences and their complement of protein motifs does however show some indication of relatedness. For example, group 5 sequences possess motif 9 (predicted by Prosite to contain a possible N-myristoylation motif). These DOFs are all expressed in root mRNA. Despite this, most of the protein motifs tentatively identified do not correlate with expression in a specific tissue.

1.2.9 - An inventory of biological roles

As previously mentioned, DOF transcription factors have a great diversity of functions and play crucial roles in many physiological processes. DOFs tend to be associated with plant-specific processes such as light-responsiveness, seed development, or germination, consistent with their origin after the divergence of the plant kingdom branch. The processes which are known to involve regulation by DOF transcription factors are summarized in Table I-3, divided into three broad categories, those having developmental roles such as seed development or tissue differentiation and those involved in regulation of metabolism. It should be noted that some DOFs have roles which overlap these categories, and that for many DOFs, no evidence of their biological role is available. However, it is apparent that DOFs do not play roles in the initial determination of organs via the resolution of meristems, in contrast to MADS-box or homeobox transcription factor classes, for example. This may reflect their more recent plant-restricted ancestry.

There are efforts underway to identify DOF gene functions systematically. For example, the over-expression in transgenic plants of the complete set (36) of *A. thaliana* DOF TFs has been undertaken (Tsujimoto-Inui et al., 2009). This group screened the transgenic plants obtained for changes in the metabolome and identified DOF genes putatively involved in regulation of enzymes of the phenylpropanoid and flavonoid pathways. Using a different strategy, GmDof4 and GmDof1 were selected from a set of soybean DOF genes for their expression during seed development, and overexpressed in *A. thaliana* plants, where they caused an increase in seed lipid content, whilst down-regulating a seed storage protein gene, *CRA1* (Wang et al., 2007).

1.2.9.1 - Tissue differentiation

ZmDOF1 is an example of many DOF factors acting on plant development. It regulates pollen-gene specific expression by repressing genes controlling pollen development (Yanagisawa, 2000; Chen et al., 2012). In support of this finding, male gamete-specific promoters, (pGEX and pGEX2), have been shown to

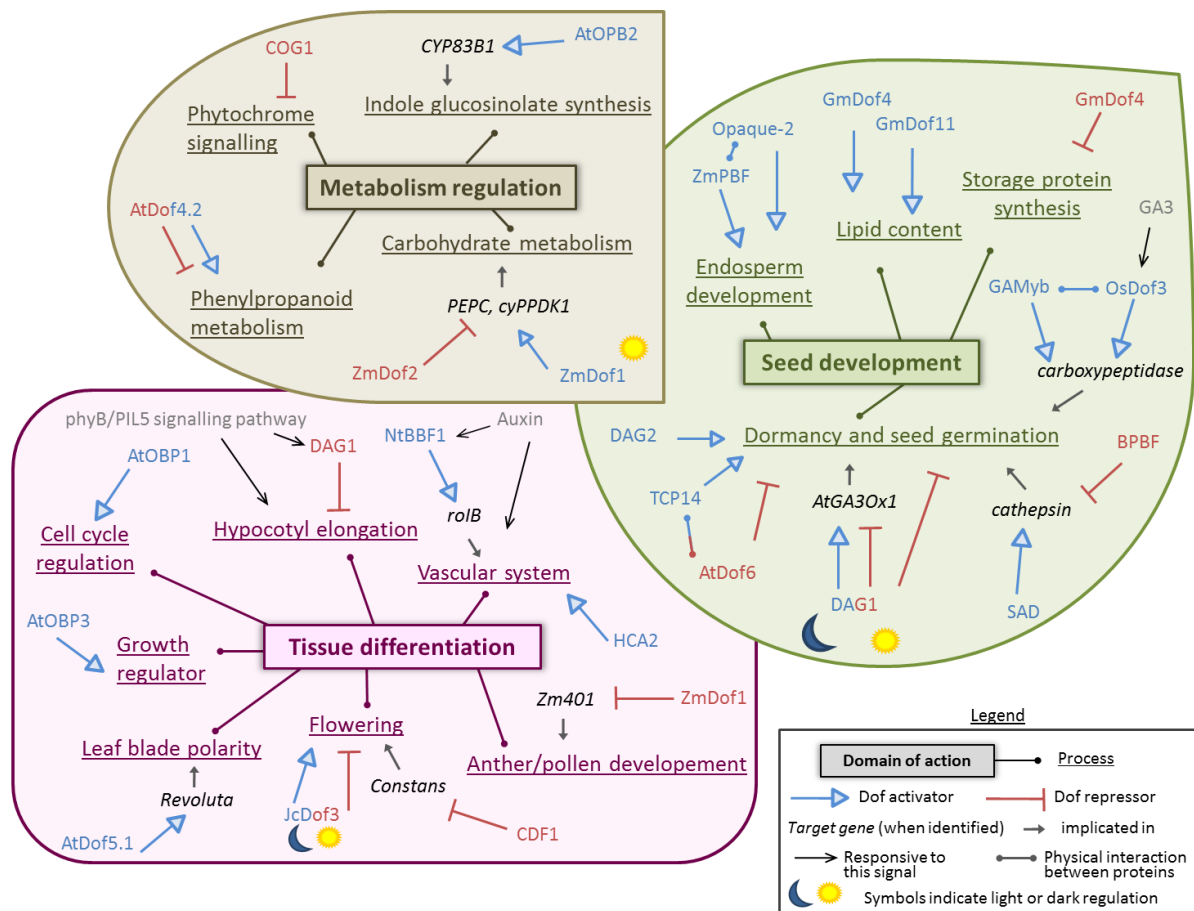


Figure I-5: Examples of some processes which are known to involve regulation by DOF transcription factors (the genes described are detailed in table II-3 and section I-2-9).

carry putative DOF binding sites (Engel et al., 2005). The *A. thaliana* DOF factor OBP1, (OBF-binding factor-1) implicated in a more general control of cell division, in promoting cell cycle re-entry (Skirycz et al., 2008). OBP1 activates the core cell cycle gene *CYCD3;3*, and a second DOF, the replication-specific transcription factor gene *AtDOF2;3*, thus shortening cell cycle duration when overexpressed.

An *A. thaliana* DOF, AtDof4.1, has been shown to be a repressor acting to regulate organ size (Ahmad et al., 2013). The DOF transcription factor *AtDOF5.1* activates expression of *Revoluta (REV)*, a class III HD-ZIP transcription factor which is required for establishment of leaf adaxial-abaxial polarity (Kim et al., 2010). *Cycling DOF Factor1 (CDF1)* represses transcription of *Constans (CO)* and thus represses flowering in *A. thaliana*. CDF1 is the target of FKF1, (Flavin-binding, Kelch repeat, F-box-1), an F-box protein that binds to it and targets it for degradation in response to light (long days) via the ubiquitin-dependent pathway, thus contributing to flowering induction (Song et al., 2012). An orthologue of *A. thaliana* CDF1 was found in potato *Solanum tuberosum L. (S. tuberosum L.)*, where it controls tuber formation by repressing *S. tuberosum L. CO1/2* in a cascade reaction leading to the induction of tuber development (Kloosterman et al., 2013). A number of other DOF factors show light-regulated expression, responding to both phytochrome and cryptochrome signals (Ward et al., 2005; Li et al., 2009), (Park et al., 2003), or are subject to circadian rhythm control (Yang et al., 2011). A dominant mutant isolated by activation tagging, *COG-1d*, revealed a negative control of phytochrome signaling by PhyA and PhyB exerted by the DOF *COG1* (Park et al., 2003). The role played by the *DOF gene Affecting Germination-1 (DAG1)* in *A. thaliana* testa differentiation is also influenced by light (Papi et al., 2002), whereas the DOF *Rice Dof daily flucutations, (Rdd1)*, is controlled by both Phy A and Phy B, and by the circadian clock (Iwamoto et al., 2009).

A. thaliana vascular development is subject to regulation by two DOF TFs, AtDof2.4 and AtDOF5.8, which are differentially expressed. AtDof2.4 is detected in procambial cells, and AtDof5.8 in cells of prospective leaf veins and in the vascular tissue of developing flower buds (Konishi and Yanagisawa, 2007). The vascular system is also the site of expression of genes involved in lignin production which possess DOF binding sites in their promoters (Rogers et al., 2005). In companion cells, both DOF and HD-ZIP TF binding sites seem important for cell-specific expression of the *SUCROSE TRANSPORTER 2* gene (Schneidereit et al., 2008). Numerous publications implicate DOF TFs in the regulation of guard cell development and function. The promoter of a myrosinase gene, which contained several DOF binding sites, was specifically expressed guard cells and idioblasts (Thangstad et al., 2004). The role of DOF factors in regulating guard cell-specific expression of the Myb60 transcription factor has been demonstrated by deletion and resection of putative DOF binding sites in its promoter, and a candidate DOF has recently been identified (Cominelli et al., 2011). StDOF1 was also identified as a candidate for activation of guard cell-specific expression. It interacts with and activates expression of a guard cell-specific gene, *KST1* (Plesch et al., 2001). An *A. thaliana* DOF TF gene isolated by positional cloning of a stomatal development mutant, *stomatal carpenter-1, (SCAP1)* is expressed in maturing guard cells, but not in guard mother cells. *SCAP1* regulates the expression of genes

involved in stomatal maturation and functioning, such as K⁺ channel protein, MYB60 transcription factor, and pectin methylesterase (Negi et al., 2013).

AtDOF4.7 is expressed in floral organ abscission zones, where it represses the expression of a polygalacturonase gene *PGAZAT*, putatively important for floral organ abscission. Interestingly, AtDOF4.7 interacts with a second Zn finger TF, AtZFP2, also implicated in repression of floral organ abscission, suggesting cooperativity between the two TFs (Wei et al., 2010).

Several DOF factors are implicated in phytohormone action. NtBBF1 binds to the promoter of the *rolB* oncogene in an auxin-regulated manner (Paolis et al., 1996), and thus promotes root growth (Baumann et al., 1999). *A. thaliana* OBP3 (AtDof3.6) expression is also under auxin control and has been ascribed a role as a regulator of plant growth based on the phenotype of its over-expression in transgenic *A. thaliana* (Kang and Singh, 2000). A DOF protein binding to the promoter of the pumpkin ascorbate oxidase gene (*AOBP*), is highly expressed in growing tissues, and is also auxin-induced (Esaka et al., 1992; Kisu et al., 1998). OBP3 (AtDof3.6) is also induced by salicylic acid (Kang et al., 2003). A series of putative target genes of OBP3 were identified by studying transcripts affected by overexpressing the gene and by inactivating it by RNAi (Kang et al., 2003). These OBP3-responsive genes (*ORGs*) included two sequences related to bHLH transcription factors and an extensin. *ORG1* to -3 are upregulated by salicylic acid and downregulated by jasmonic acid.

1.2.9.2 - Seed development

Several DOF transcription factors are implicated in the regulation of synthesis of storage proteins and other proteins expressed during the seed-filling phase of endosperm development in cereals. They recognize a TGTAAG motif present in the Prolamin-box on the corresponding promoters, and are termed Prolamin Box Binding Factors (PBFs). PBFs sharing a high degree of sequence similarity have been described in maize ZmPBF (Vicente-Carbajosa et al., 1997; Marzabal et al., 2008), rice PBF (Yamamoto et al., 2006), wheat WPBF (Mena et al., 1998; Dong et al., 2007), barley BPBF (Mena et al., 1998) and finger millet FMPBF where they have similar expression profiles, being restricted to the developing endosperm and appearing at the beginning of the seed filling phase. They interact with Opaque-2 type bZIP TFs, and have been shown to contribute to the activation of transcription of storage protein target promoters during seed development (Vicente-Carbajosa et al., 1997; Marzabal et al., 2008). Besides storage protein genes, other endosperm-expressed genes are regulated by the combination of PBFs and Opaque-2-like bZIPs, such as lysine ketoglutarate reductase/saccharopine dehydrogenase (LKR). However, the mechanism of regulation may be different in this case, as the presence of either the PBF or the bZIP suffices to maintain storage protein gene expression, whereas LKR expression requires both factors (Kawakatsu and Takaiwa, 2010). A further study of rice, *A. thaliana* and wheat DOF genes revealed that several DOF genes other than the putative PBF orthologs are also expressed during seed filling and that these are likely to have different roles (Gaur et al., 2011).

Further DOF genes have pivotal roles in controlling seed germination. A DOF gene affecting germination, but expressed in the maternal testa layer of the seed, (*DOF affecting germination, DAG1*), reduce dormancy when mutated (Papi et al., 2000). A second DOF gene, (*DAG2*), although a closely related paralog, showed the opposite effect, ie., increased dormancy and reduced response to gibberellic acid when mutated (Gualberti et al., 2002). It seems these two genes act antagonistically on the same set of target promoters, *DAG1* repressing and *DAG2* activating, the *dag1* mutation being epistatic over *dag2*. *DAG1* mediates the control of germination by light, by repressing *GA3ox1* expression and hence GA4 formation (Gabriele et al., 2010). *DAG1* is activated indirectly in the dark by the bHLH transcription factor *PIL5* (*phytochrome interacting factor 3 like-5*), which responds to phytochrome B. In the *dag1* mutant, GA4 synthesis is de-repressed and germination is facilitated. *DAG1* may act partly by affecting chloroplast development during seed maturation. For example, it appears to control indirectly the expression of the chloroplast-located early light-induced proteins, ELIP1 and ELIP2, themselves needed for seed germination (Rizza et al., 2011). A DOF gene activating cathepsin expression during germination in the barley aleurone, *SAD*, was identified (Isabel-LaMoneda et al., 2003). Further work revealed the involvement of other transcription factors in regulating cathepsin gene expression, including a second DOF, barley prolamin box binding factor (BPBF), which acts antagonistically to *SAD*, and represses cathepsin expression during seed development (Zou et al., 2008). A further gene target of BPBF, *Alpha-amylase 32b* is regulated similarly to cathepsin (Zou et al., 2008), being repressed during seed development and activated during seed germination. The *Alpha-amylase 32b* promoter bears a cluster of 4 binding sites (W-box-Pyr-GARE-Amy) to which binds a “repressosome” consisting of WRK38, the DOF BPBF, HRT, (a GARE-binding protein) and the Myb MCB1 (Zou et al., 2008). The repressosome is replaced during germination by an “enhanceosome” in which these four transcription factors are substituted by different members of the same classes of transcription factors: RAMY, SAD, GAMYB, and MYBS3. Physical interactions of the DOFs with transcription factors binding at adjacent positions on the promoter were shown by bimolecular fluorescence complementation. In addition to the interactions reported for the repressosome and enhanceosome, PBFs interact with bZIP TFs (Vicente-Carbajosa et al., 1997), and with TaQM, a seed protein possessing a sequence closely related to that of a ribosomal protein, during seed development (Dong et al., 2007).

Rice OsDOF3 protein, which is expressed in germinating seeds, binds the pyrimidine boxes (AAAG motifs) located in the *CPD3* (type3 carboxypeptidase) promoter, and enhances its expression under gibberellin control (Washio, 2001, , 2003). A further DOF negative regulator of seed germination, *AtDOF6*, putatively exerts its effect by interacting with *TCP14*, a positive regulator of seed germination (Rueda-Romero et al., 2012). TCP-domain proteins from rice and *A. thaliana* bind to the site II motif, a *cis*-element directing cell cycle- and protein synthesis-related gene expression (Kosugi et al., 1995; Kosugi and Ohashi, 1997; Tremousaygue et al., 2003), suggesting a possible inhibitory action of *AtDOF6* on these processes.

I.2.9.3 - Regulation of metabolism

DOF factors intervene in many aspects of metabolic responses to environmental inputs to control plant growth responses. Members of the DOF protein family are involved in light-regulation of genes involved in carbohydrate metabolism (Yanagisawa and Sheen, 1998). Maize DOF1 enhances transcription from the maize C4 phosphoenol pyruvate carboxylase promoter, whereas DOF2 blocks this transactivation and represses C4 PEP carboxylase expression, although it is able to transactivate other promoters (Yanagisawa and Sheen, 1998; Yanagisawa, 2000). The repression by DOF2 in non-leaf tissues is speculated to restrict C4 PEPC carboxylase expression to leaves, where DOF2 is not expressed. The positive effect of Dof1 on carbon skeleton production has been exploited for metabolic engineering of *A. thaliana* (Yanagisawa et al., 2004). By over-expressing Dof1, phosphoenol pyruvate carboxylase and pyruvate kinase activities were elevated, and the transgenic plants exhibited enhanced nitrogen assimilation, presumably limited by C-skeleton availability. The transgenic plants were also able to grow better under low-N conditions, a finding also reported for rice transgenics expressing the ZmDof1 gene (Kurai et al., 2011). Interestingly, reducing ZmDof1 transcript levels by 80% did not affect C4-specific gene expression in maize, suggesting the possibility of redundancy in its action on these promoters (Cavalar et al., 2007). In sweet potato (*Ipomoea batatas*), over-expression of a storage root-factor DOF gene, *SRF1*, modified carbohydrate metabolism of the tuber, notably by reducing vacuolar invertase accumulation to 30% via diminution of the *lb6fruct2* transcript abundance (Tanaka et al., 2009). The effect was to increase starch content at the expense of reduced monosaccharide concentrations.

Overexpression of OsDof25 in *A. thaliana* altered carbon and nitrogen metabolism to result in increases in amino acid concentrations (Santos et al., 2012). *OsDof25* increased the levels of expression of high and low affinity ammonium transporters (*AtAMT1.1* and *AtAMT2.1*, respectively) and repressed the high affinity nitrate transporter (*AtNRT2.1*). An increase in total amino-N content and expressions of the pyruvate kinase (*PK1* and *PK2*), phosphoenolpyruvate carboxylase (*PEPC1* and *PEPC2*) and NADP-dependent and NAD-dependent isocitrate dehydrogenase was seen and in addition, an increase in expression levels and activity of glutamate dehydrogenase was observed. The simultaneous increase in the expression of *AMTs* and enzymes of carbon metabolism may explain the high levels of amino-N found in the transgenic plants.

A DOF transcription factor gene, *PpDof5*, expressed in maritime pine (*Pinus pinaster*) is implicated in the regulation of two glutamine synthase genes, *GS1a* and *GS1b* (Rueda-Lopez et al., 2008). *PpDof5* contributes to the activation of the promoter of *GS1b* and to the repression that of *GS1a* in GUS reporter gene assays conducted by transfecting pine protoplasts, and is proposed to play a role in nitrogen reassimilation during lignification.

By comparing a global microarray analysis of *Physcomitrella patens* gene expression with that of *A. thaliana* under various physiological conditions, it could be deduced that the regulation of certain functions, such as salt stress responses, involves the same families of TFs, including DOFs in the moss and in embryophytes, whereas other programs controlling for example seed and flower development in embryophytes have been elaborated from light signalling and ABA response regulatory circuits (Richardt et al., 2010). This observation may help to explain the different spheres of activity occupied by the different families of TFs in higher plants. In *P. patens*, PpDof1 is implicated in the control of caulonema filament growth. PpDof1 appears to promote filament growth under conditions of high carbon:nitrogen ratio, and is therefore dependent on environmental conditions. Both PpDof1 and PpDof2 behave as repressors in maize protoplast transient expression assays, and the double mutant created by homologous gene replacement shows modified expression of a series of putative target genes (Sugiyama et al., 2012).

DOF genes are important modulators of plants responses to both abiotic and biotic stresses. In this context, many are induced by different phytohormones and/or by pathogens, and participate in the regulation of phytoalexins (Nakano et al., 2006; Skirycz et al., 2006; Skirycz et al., 2007). Furthermore, promoters responding to pathogen-related elicitors often contain putative DOF TF binding sites (Gomez-Ros et al., 2012), (Haque et al., 2008), (Kang et al., 2003), (Larsen, 2004), (Mondal et al., 2007), (Zheng et al., 2012). There are however relatively few examples of functional validation of the roles played by DOF TFs in pathogen defense. OBP2 is implicated in indole glucosinolate biosynthesis in *A. thaliana*. It is vascular tissue-specific and induced in response to a generalist herbivore, *Spodoptera littoralis*, and by methyl jasmonate, both of which also trigger indole glucosinolate accumulation. By using AtDOF4.2 overexpressing and antisense plants, it was shown that AtDOF4.2 affects phenylpropanoid metabolism in *A. thaliana*, negatively affecting flavonoid synthesis, but positively influencing hydrocinnamic acid synthesis (Skirycz et al., 2007).

20 years since the discovery of the first DOF transcription factors, their involvement in numerous aspects of plant development has been demonstrated, but the direct gene targets have not been unambiguously identified in most cases, and much remains to be understood about the regulation of their action, their interactions with other transcription factors, their post-transcriptional regulation and turnover. We need a better appreciation of why this particular class of transcription factors has been recruited for the roles attributed to it. Among the priorities to be addressed are to overcome the common problem of functional redundancy of DOF genes by obtaining multiple gene knockouts and thereby be able to assign functions to these genes, to determine the roles in transcription of the sequences flanking the DOF domain, and to test the role if any in regulation of the non-zinc finger Cysteine residue within the DOF domain.

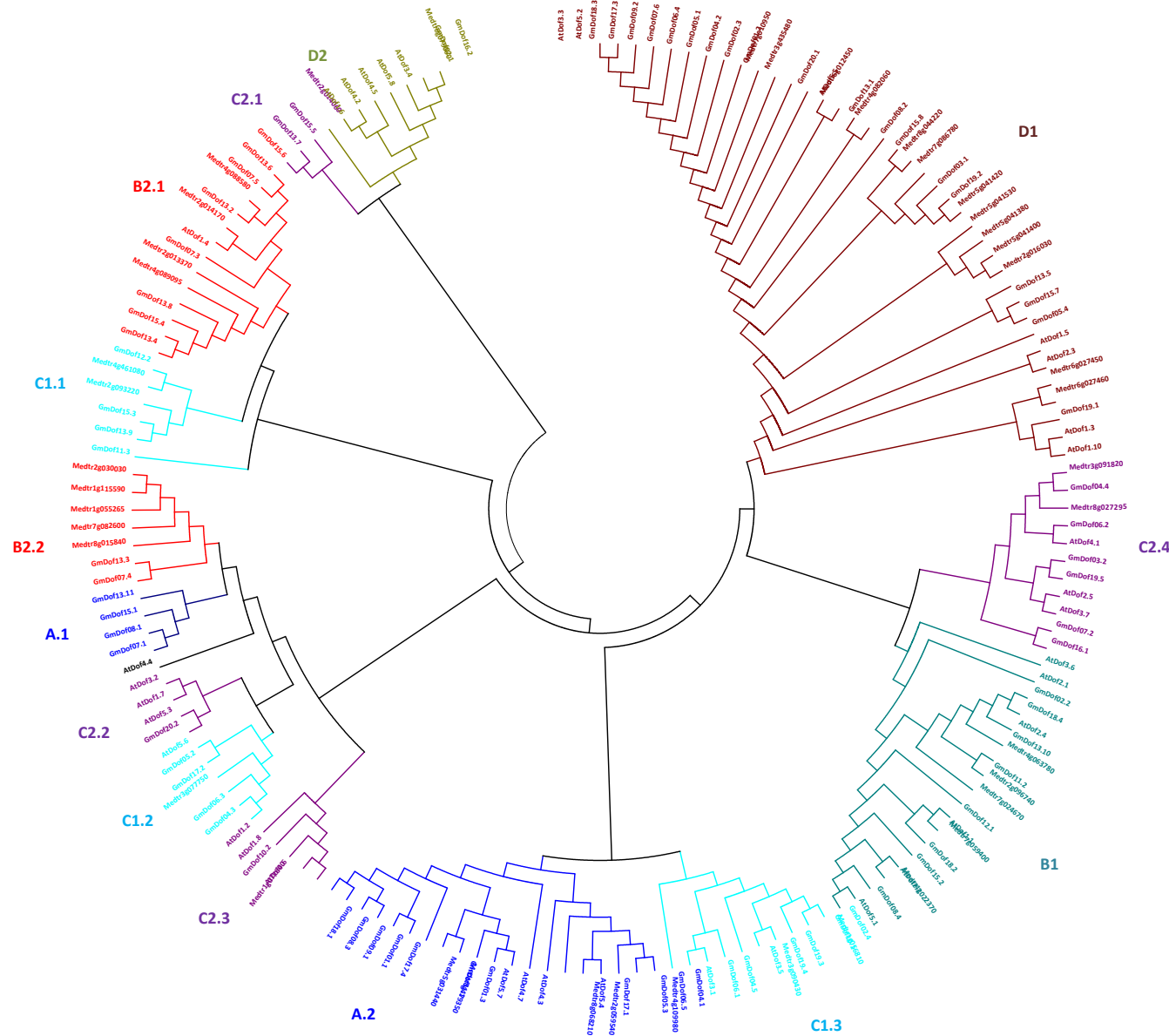


Figure I-6 – Arbre phylogénique représentatif des différentes séquences protéiques DOF chez *A. thaliana*, *G. max* et *M. truncatula*. L'alignement des 36 séquences d'*A. thaliana*, 78 de *G. max* et 42 de *M. truncatula* a été réalisé par ClustalW et l'arbre via MEGA avec la méthode neighbor-joining, les sous-groupes identifiés d'après les groupes définis par Guo and Qiu (2013) sont indiqués par un code couleur similaire.

Acknowledgements

We thank the anonymous reviewers for their suggested improvements to the manuscript. Work in the authors' laboratory is supported by the European Commission (EU-FP7 ABSTRESS). MN is supported by a bursary co-financed by INRA DGAP and the Conseil Régional de Bourgogne. RMA was supported by a studentship from the Higher Education Council of the Government of Pakistan.

1.3 - Conclusion

Cette revue donne un aperçu de la diversité de cette famille de FT de type DOF au sein de plusieurs espèces. L'étude phylogénique a montré qu'il est cependant très difficile de trouver un ancêtre commun à cette famille tant les séquences bordant le domaine DOF sont variables. Il apparaît ainsi plus simple de restreindre l'étude comparative des séquences de DOF chez deux ou trois espèces comme cela a été fait à plusieurs reprises : *A. thaliana* avec le riz (Lijavetzky et al., 2003), *Brachypodium distachyon* avec le riz et l'orge (Hernando-Amado et al., 2012), *A. thaliana* avec le sorgho et le riz (Kushwaha et al., 2011), *A. thaliana* avec le riz, la mousse et l'unique séquence retrouvée chez *Chlamydomonas* (Shigyo et al., 2007). Ces arbres permettent ainsi une meilleure interprétation des séquences homologues.

Une étude récente des FTs de type DOF chez le soja dénombre 78 séquences et l'arbre phylogénique regroupant ces 78 séquences avec les 36 séquences d'*A. thaliana* et les 30 du riz y est présenté: neuf groupes ont été définis sur la base d'une homologie de séquences [annexe I-1] (Guo and Qiu, 2013). Nous avons utilisé cette même approche avec ces 78 séquences de soja, 36 d'*A. thaliana* et 42 de *M. truncatula* annotées comme des séquences de type DOF dans la nouvelle version du génome 4.0 (<http://www.jcvi.org/medicago/>). L'arbre phylogénique a été réalisé grâce aux mêmes logiciels: alignement des séquences protéiques via ClustalW et visualisation de l'arbre par MEGA (Tamura et al., 2013) avec la méthode Neighbor-joining. La figure I-6 présente cet arbre et le code couleur reprend celui de Guo et al. (Guo and Qiu, 2013), basé sur les séquences de soja. Les clusters B1, D1 et D2 sont identiques à ceux précédemment décrits dans les travaux de Guo and Qiu (2013). Les autres sont éclatés en deux ou trois groupes: A, réparti en deux sous-groupes A.1 et A.2 ; B2, divisé en B2.1 et B2.2 ; C1, découpé en C1.1, C1.2 et C1.3 et enfin C2, subdivisé en quatre, de C2.1 à C2.4. Les séquences d'*A. thaliana* sont quant à elle réparties de façon différentes dans les groupes, ce qui confirme que l'étude de la phylogénie de ces séquences reste difficiles à interpréter étant donné la variabilité de séquences de part et d'autre du domaine dof.

Les séquences DOF de *M. truncatula* pour lesquelles des séquences DOF orthologues chez le soja ont été identifiées d'après l'arbre sont présentées dans le tableau I 4. Les sites d'expression de ces gènes dans la plante et plus spécifiquement dans la graine ont été recherchés, de même que les données cinétiques au

CHAPITRE -1-

	Medicago truncatula																			Glycine max												
groupe [figure I-6]	identifiant v4.0	identifiant Mt3.5 (>85% de résidues identiques à v4.0)	site d'expression dans la graine à 12 jap (log2)*			Cinétique d'expression au cours du développement de la graine														autre site d'expression dans les tissus de la plante *	séquence proche d'après l'arbre [figure I-6]			site d'expression dans la graine*	Cinétique d'expression au cours du développement de la graine							autre site d'expression dans les tissus de la plante *
			Alb	Teg	Emb	8 jap	11 jap	14 jap	17 jap	20 jap	23 jap	26 jap	29 jap	32 jap	35 jap	38 jap	41 jap	44 jap	10 jap						14 jap	21 jap	25 jap	28 jap	35 jap	42 jap		
D2	Medtr8g079060.1	IMGA contig_9183_2.1	0,2	0,2	3,9	2,7	3,1	3,3	2,0	1,4	1,6	1,6	1,6	2,1	1,6	1,9	1,8	1,4	ND	Glyma02g06970.1	GmDof2.1	ND	1	9	7	4	3			-		
																				Glyma16g26030.1	GmDof16.2	ND	3	5	4	2			gousse			
C2.1	Medtr2g014060.1	Medtr2g014060.1	4,4	2,7	1,8	3,8	5,1	4,7	1,7									1,9	-	Glyma13g31110.1	GmDof13.7	ND								racines, jeunes feuilles (faible)		
																				Glyma15g08230.1	GmDof15.5	ND			1				-			
B2.1	Medtr4g088580.1	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	Glyma07g31870.2	GmDof7.5	emb (GS, HS, CS)	3	2	2	1		1	1	-		
	Medtr2g014170.1	Medtr2g014170.1	0,1	2,1	2,1	2,6	2,3	2,5	2,9	2,2	1,9	1,5	1,7	1,8	1,3				nodules; fleurs	Glyma13g24600.1	GmDof13.2	emb (CS)	4	2	4	2				jeunes feuilles		
C1.1	Medtr4g461080.1	contig.17531.1.1	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	Glyma12g07710.1	GmDof12.2	teg externe (GS, HS, CS); emb (CS), teg FVC (MS)	7	16	16	7		7		gosses (10 jap)		
	Medtr2g093220.1	Medtr2g093220.1	1,3	3,4	5,2	5,1	4,6	4,3	4,7	3,4	3,5	3,7	2,6	3,3	2,7		3,4		gosses; racines	Glyma13g40420.1	GmDof13.9	teg interne (GS, HS, CS, MS)	2	1	1	1	1	2	1	gosses (14 jap); racines		
																				Glyma15g04980.2	GmDof15.3			2	1	2		1		gosses, jeunes feuilles (faible)		
C1.2	Medtr3g077750.1	Medtr3g077750.1	0,6	1,4	2,5	4,4	3,0	2,5	4,7	3,7	3,4	2,2	1,6	2,0					pétioles	Glyma04g35650.1	GmDof4.3	pas d'expression						2	3	gosses (10 et 14 jap); racines; jeunes feuilles		
																				Glyma06g19330.1	GmDof6.3	pas d'expression		2	2				2	gosses 10 et 14 jap (faible)		
C2.3	Medtr1g077600.1	IMGA contig_76558.1. 1	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	Glyma10g31700.1	GmDof10.2	pas d'expression	1	2	1	1		1		gosses 10 et 14 jap (faible)		
A.2	Medtr5g031440.1	Medtr5g031440.1	5,9	2,6	2,9		3,6	4,2	2,9	2,0	1,5			1,7					ND	Glyma17g21540.1	GmDof17.4	ND	6	6	11					-		
	Medtr4g109980.1	IMGA contig_52073.1. 1	1,3	3,1	1,5	3,9	3,7	4,0	2,2	1,4	1,1			1,9					ND	Glyma17g08950.1	GmDof17.1	ND	1	3		5	3	4	3	-		
C1.3	Medtr3g090430.1	Medtr3g090430.1	0,6	0,5	0,3	4,0			1,3	1,1				1,6	1,6	1,7	1,1		racines	Glyma06g12950.1	GmDof6.1	teg interne (GS, HS), teg (CS, MS)	1	3	2	4	3	3	1	graines (25, 28 et 35 jap)		
																				Glyma04g41830.2	GmDof4.5	ND								-		
	Medtr1g056810.1	IMGA contig_239421. 1.1	0,2	0,4	0,5		0,4	0,8	0,6	0,6	0,5	0,3	0,3	0,5			0,1	0,4	ND	Glyma10g10142.1	GmDof10.1	ND	ND	ND	ND	ND	ND	ND	ND	ND		
																				Glyma02g35296.1	GmDof2.4	ND	ND	ND	ND	ND	ND	ND	ND	ND		
B1	Medtr4g022370.1	Medtr4g022370.1	ND	ND	ND	0,3	0,4	0,7	0,7	0,3	0,3			0,3	0,2	0,2	0,3	0,3	racines; nodules	Glyma08g47290.1	GmDof8.4	ND	2	2	4	2		3		jeunes feuilles		
	Medtr7g059400.1	IMGA contig_99303.1. 1	0,7	2,4	2,5	2,1	2,4	2,1	3,1	2,5	2,8	1,9		2,9					ND	Glyma12g06880.1	GmDof12.1	ND	1		2			1	1	-		
	Medtr4g063780.1	IMGA contig_49316.1. 1	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	Glyma11g14920.2	GmDof11.2	ND	1	1		1	1	1	1		racine faible)	
C2.4	Medtr3g091820.1	ND	ND	ND	ND	5,5	4,7	4,2	4,5	4,0	3,5			4,2	4,1	4,0	4,2		ND	Glyma04g41170.1	GmDof4.4	ND	3	2	2	1		1	1	-		
D1	Medtr2g016030.1	IMGA contig.15234.1. 1	1,1	3,3	1,1	4,0			4,4	3,8	3,6	3,8	3,8	4,1	5,1	5,2	5,4	4,0	ND	Glyma15g08860.1	GmDof15.7	ND								-		
																				Glyma13g30331.1	GmDof13.5	ND	ND	ND	ND	ND	ND	ND	ND	ND		
	Medtr7g086780.1	Medtr7g086780.1	3,9	3,2	1,3	3,9	3,9	3,2	4,1					3,9	4,8	5,0	6,3	5,3	-	Glyma03g01030.1	GmDof3.1	ND		1		1		1	2	nodules		
																				Glyma19g29610.2	GmDof19.2	ND		1						-		
	Medtr6g012450.1	Medtr6g012450.1	3,9	4,2	4,1			4,6	5,6	4,3	4,6			4,6	4,6	4,7	4,6		-	Glyma13g05480.1	GmDof13.1	ND		1		1		2	1	-		

*site d'expression dans les tissus de la graine d'après des données d'hybridation Nimblegen (cf chapitre 3); cinétique au cours du développement de la graine (données non publiées – K. Righetti, J. Bultink – UMR PMS Angers) et site d'expression dans les tissus de la plante déterminé d'après Mt Gene Expression Atlas (<http://mtgea.noble.org/v3/>) (Benedito et al., 2008; He et al., 2009).

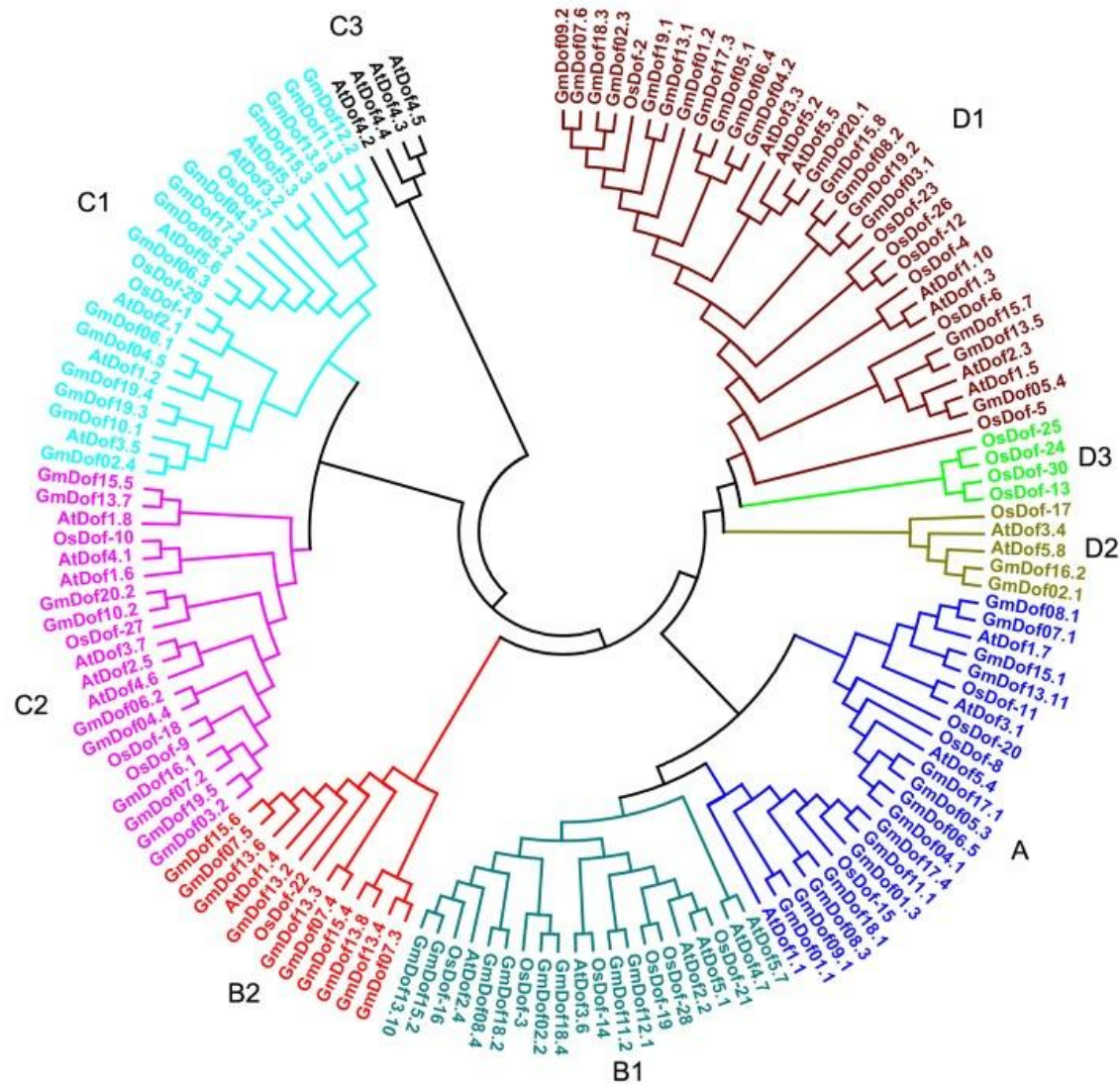
*site d'expression dans les tissus de la graine d'après les données disponibles sur <http://seedgenenetwork.net/soybean> et dans les tissus et cinétique au cours du développement de la graine déterminés d'après soyseq (<http://soybase.org/soyseq/>; données normalisées, Severin et al., 2010)

Tableau I-4 - Gènes *DOF* chez *M. truncatula* dont des séquences proches ont été retrouvés chez le soja (d'après l'arbre présenté dans la figure I-6). Leur site d'expression dans les tissus de la plante et de la graine ainsi que leur cinétique d'expression au cours du développement de la graine sont indiqués. alb: albumen, emb: embryon, teg:tégument, GS: stade globulaire, HS: stade cœur, CS: stade cotylédons, MS: stade mature, JAP: jours après pollinisation, FVC: faisceaux vasculaires des cotylédons.

cours du développement de la graine (*M. truncatula* d'après des données non publiées – K. Righetti, J. Buitink – UMR IRHS Angers; *G. max* d'après les portails <http://seedgenenetwork.net/soybean> et <http://soybase.org/soyseq/>, (Severin et al., 2010)). On retrouve des similitudes de profils d'expression pour plusieurs séquences: dans le groupe D2, où Medtr8g079060.1 et GmDof2.1 ont une expression maximale vers 14 jap, ou encore dans le groupe B2.1 où Medtr2g014170.1, GmDof5.7 et GmDof13.2 ont une expression préférentielle en début de développement de la graine. On observe également que ces trois gènes sont exprimés dans l'embryon. La séquence Medtr7g086780.1, présente au sein du groupe D1, est proche de GmDof3.1 et ces deux gènes partagent une expression plus forte en fin de maturation, vers 42 jap. La recherche de ces informations nous a permis de constater pour quelques gènes des profils d'expression dans les tissus de la graine et/ou au cours du développement concordants entre les deux espèces *M. truncatula* et *G. max*. Les mécanismes de régulation impliquant ces FTs de type DOF seraient donc conservés entre espèces légumineuses. Toutefois dans la majorité des cas bien que ces séquences soient proches en termes d'homologie, elles ne codent pas systématiquement pour des FTs s'exprimant dans les mêmes tissus ou aux même stades de développement de la graine.

Les FTs de type DOF sont impliqués dans des processus physiologiques divers comme la germination (Papi et al., 2000; Gualberti et al., 2002; Gabriele et al., 2010), les mécanismes de défense envers les pathogènes (Chen et al., 1996) ou encore la régulation du développement (Park et al., 2003; Yang et al., 2011; Song et al., 2012; Kloosterman et al., 2013). Ils sont également retrouvés dans la régulation de la synthèse des protéines de stockage ou des protéines spécifiques de l'albumen chez les céréales (Vicente-Carbajosa et al., 1997; Mena et al., 1998; Yamamoto et al., 2006; Dong et al., 2007; Marzabal et al., 2008). Cependant, aucun gène DOF n'a été identifié comme ayant un rôle dans le développement de la graine au niveau de l'embryogénèse chez les dicotylédones. Parmi les gènes DOF s'exprimant plus fortement dans la graine par rapport aux autres tissus de la plante [figure I-3 et tableau I-4], on note la présence du gène Medtr2g014060.1, fortement exprimé en début de développement de la graine vers 11-14 jap. Le gène DOF15.5 identifié chez le soja est un orthologue possible de ce gène. Par ailleurs, des données RNAseq non publiées (obtenues chez le pois par l'équipe de Judith Burstin) nous ont permis d'identifier un gène présentant une forte homologie de séquences avec Medtr2g014060.1 et exprimé au même stade du développement de la graine de pois. Ces données suggèrent une conservation de sa fonction entre espèces légumineuses. Ce gène présente donc un intérêt particulier de par son expression au cours de l'embryogénèse et son fort niveau d'expression dans l'albumen des graines. La description de ce gène Medtr2g014060.1, renommé *DASH* pour DOF Acting in Seed embryogenesis and Hormone regulation, fera l'objet du chapitre II (article en préparation).

I.4 - Annexes



Annexe I-1 – Arbre phylogénique extrait de l'article de Guo et al. 2013.

Chapitre II

-

*Etude d'un facteur de transcription
de type DOF impliqué dans le
développement de la graine chez
Medicago truncatula*

II.1 - Introduction générale

La recherche des gènes impliqués dans le contrôle des critères relatifs à la qualité de la graine (composition protéique, taille des graines) peut être abordée par différentes approches. L'une d'entre elles consiste à rechercher les régions génomiques impliquées dans le déterminisme de ces caractères: détection de QTL (Quantitative Trait Loci) puis à cartographier plus finement des QTL d'intérêt et cloner les gènes sous-jacents impliqués. Une autre stratégie vise à sélectionner les gènes exprimés à des stades donnés du développement de la graine associés à l'embryogénèse ou au remplissage, puis à étudier leur colocalisation éventuelle avec des QTL de critères de qualité des graines. La fonction des gènes candidats positionnels (*i.e.* sous-jacents aux QTL) est ensuite validée au moyen de mutants. Cette dernière stratégie a été utilisée afin de caractériser la fonction d'un facteur de transcription (FT) de type DOF (DNA binding with one finger), dont le gène est sous-jacent à un QTL de poids de mille grains (données non publiées issues du projet ANR QualityLegSeed 2007-2010) et préférentiellement exprimé dans la graine pendant l'embryogénèse. Ce FT avait également été identifié par une étude des FTs exprimés pendant le remplissage de la graine (Verdier et al., 2008). La caractérisation de ce FT grâce à l'étude de mutants (EMS et Tnt) pour ce gène a permis de décrire finement le phénotype associé à ces mutations. Une approche comparative du transcriptomique des gousses d'un de ces mutants vs WT a également été conduite afin de mettre en évidence les voies métaboliques affectées par la mutation.

II.2 - DASH, un régulateur de l'embryogenèse (article en préparation)

DASH transcription factor regulates seed size and embryo morphogenesis by disturbing auxin homeostasis

Mélanie Noguero^{*1}, Christine Le Signor^{*1}, Vanessa Vernoud¹, Kaustav Bandyopadhyay², Myriam Sanchez¹, Chunxiang Fu², Ivone Torres-Jerez², Julia Buitink⁴, Karine Gallardo¹, Michael Udvardi², Richard Thompson¹ and Jerome Verdier^{2, 3}

¹INRA, UMR Agroécologie, pôle GEAPSI, Dijon, FR

²The Samuel Roberts Noble Foundation, Plant Biology Division, Ardmore, OK, USA

³Shanghai Center for Plant Stress Biology, Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences, Shanghai, PR China

⁴ UMR 1191 Physiologie Moléculaire des Semences, Université d'Angers/INH/INRA, Angers, FR

*joint first authors

II.2.1 - Introduction

Agriculture is facing the challenge of maintaining or enhancing production to match globally increasing demands, whilst reducing farm inputs and associated environmental pollution. Legumes can play a crucial role, providing protein-rich food sources whilst reducing the need for nitrogen fertilizer by their ability to fix atmospheric nitrogen. Knowledge of the factors controlling seed developmental processes in legumes might help to increase or maintain seed production and high seed quality under increasingly challenging environmental conditions. During seed development, the tissues surrounding the developing embryo (*i.e.* endosperm and seed coat) play important roles in determining both final seed size and composition, as shown by mutant phenotypes in *Arabidopsis*, maize, *Medicago truncatula* and pea (Luo et al., 2005; Adamski et al., 2009; D'Erfurth et al., 2012), and also by maternal effects on seed size in several crops, including pea (Autran et al., 2011; Li et al., 2011; Fang et al., 2012). Like in most angiosperms, after fertilization, the endosperm development is divided in two phases: a syncytial phase with nuclear divisions without cytokinesis and a cellularization phase, which will differentiate into three distinct domains: the micropylar region (or embryo surrounding region in cereals) close to the embryo, the central region (or starchy endosperm in cereals) forming the largest part of endosperm and the chalazal region (or transfer layer in cereals) involved in the transfer of maternal nutrients to embryo and/or endosperm.

In contrast to cereals, the endosperm of most dicotyledonous seeds including legumes is a transient structure, undergoing cell breakdown during seed filling to leave only a single layer of endosperm cells at maturity. However, whilst being a non-storage tissue at maturity, the endosperm has an important role to play in nourishing the developing embryo (Hehenberger et al., 2012) and thus indirectly affects final seed content. Embryo development is dependent on nutrient supply from the maternal tissues, which in cereals are channelled initially via the suspensor and subsequently mainly through the Embryo Surrounding Region (ESR) of the developing endosperm (Ingram, 2010). This zone is also thought to produce signalling molecules, which directly control embryo development (Cossegal et al., 2008). Interestingly, the endosperm has not only been implicated in the regulation of embryo development, but also in the development of the maternally derived testa (*i.e.* seed coat) (Luo et al., 2005). Final seed size has also been shown to depend on a complex cross talk between the endosperm and the integuments during early post-fertilization seed expansion (Garcia et al., 2005).

Several genes have been showed to be essential for proper development of both endosperm and embryo. For instance, HAIKU2 (IKU) is a leucine-rich repeat receptor-like kinase (LRR-RLK) expressed specifically in endosperm and *iku* mutants displayed precocious endosperm cellularization and decrease in embryo cell size, resulting in smaller seeds (Luo et al., 2005). A recent study showed that the *Arabidopsis* peptide ligand CLAVATA3/EMBRYO SURROUNDING REGION 8 (CLE8), which is restricted in its expression to the young embryo and endosperm, regulates both embryo cell divisions and endosperm proliferation (Fiume and Fletcher, 2012). It has been showed that regulation of endosperm cellularization, and consequently the final seed size, is modulated by genomic imprinting (*i.e.* effect of imbalance between the

ratio of paternal and maternal genome. Extra doses of maternal alleles cause a precocious cellularization and a reduced final seed size, although extra doses of paternal alleles result in larger seeds (Scott et al., 1998). This epigenetic effect consists of silencing parental genome via histone modulation and DNA methylation, which involves complex mechanisms and several regulatory genes including DNA methylases (see reviews, (Finnegan et al., 2000; Berger et al., 2006; Huh et al., 2007)). Thus, in addition to its conduction of nutrients, the endosperm appears to coordinate via signal transmission the development of the three seed compartments.

Analysis of the transcription factors (TFs) involved can help to better understand this regulation. In a previous study (Verdier et al., 2008), we identified several seed-specific genes expressed in the endosperm tissue of *Medicago truncatula*. Among them is a DNA-binding with One Finger (DOF) zinc finger protein, which belongs to this plant-specific transcription factor (TF) family. The DOF family is implicated in a wide range of processes essential for seed germination (Papi et al., 2000; Gualberti et al., 2002; Gabriele et al., 2010) and plant development (Park et al., 2003; Yang et al., 2011; Song et al., 2012; Kloosterman et al., 2013), including the establishment of plant defences against pathogens (Chen et al., 1996). The DOF TFs are characterized by a highly conserved DOF domain of 52 amino acids, essential for DNA binding (Yanagisawa and Sheen, 1998) and for protein-protein interactions (Yanagisawa, 1997; Kang and Singh, 2000). DOF TFs have been reported to participate in the regulation of the synthesis of seed storage proteins and other endosperm-specific proteins in cereals (Vicente-Carbajosa et al., 1997; Mena et al., 1998; Yamamoto et al., 2006; Dong et al., 2007; Marzabal et al., 2008). Notably, they bind to the endosperm motif, a conserved sequence present in the promoters of endosperm-specific genes, and they interact with a subfamily of bZIP TFs, which activate the same promoters, of which *Opaque-2* was the founder member. In contrast to monocot seeds, the role of DOF TFs in the endosperm of dicot seeds has not been studied so far.

Little information is available on regulation and action of plant hormones during early stages of seed development in both dicots and monocots. However, auxin is important for several major steps of embryo specification in both the apical and basal domains (Moller and Weijers, 2009). The auxin involved appears to be synthesized in the embryo apices and distributed throughout the embryo in gradients established by the PIN family of efflux carriers. Regarding the role and mode of action of plant hormones specifically in the endosperm, reports are scanty. High cytokinin concentration due to overproduction of isopentenyltransferase (IPT) caused a mosaic aleurone phenotype (Geisler-Lee and Gallie, 2005), suggesting cytokinin might inhibit aleurone cell fate determination. Other data point to a role for auxin in early endosperm differentiation in maize: accumulation of auxin due to auxin transport inhibition by N-1-Naphthylphthalamic Acid (NPA) caused multiple aleurone layers to develop (Forestan and Varotto, 2011), and an auxin (IAA)-deficient maize mutant, de-B18 has 30% less seed mass than the wild-type (WT) (Forestan et al., 2010). On the other hand, transcriptome analysis of proliferating endosperm in *Arabidopsis*

showed a under-representation of auxin-related genes but a significant enrichment in genes involved in CK signaling. More recently the *CKX2* gene encoding a cytokinin oxydase2 involved in cytokinn degradation has been shown to be a direct target of the IKU pathway controlling endosperm growth (Li et al., 2013, PNAS).

In order to improve our understanding of the role of the endosperm during early stages of seed development, we studied an endosperm-specific DOF TF, called hereafter *DASH* (DOF Acting in Seed embryogenesis and Hormone regulation). The *DASH* gene is specifically expressed in the developing *Medicago truncatula* seed at the phase of transition between embryogenesis and seed filling (Verdier et al., 2008). By analysing the phenotype of dash mutants, we provide evidence for a crucial role of the developing endosperm in early legume seed development. Cytological analyses demonstrated that dash mutation severely affects embryo morphogenesis, and a comparison of the transcriptomes of loss-of-function mutants and gain-of-function transformants highlighted severe changes in molecular processes related to early stages of seed development, some of which were linked to genomic imprinting and auxin signaling and transport. Most notably, dash seeds accumulate high concentrations of free and conjugated auxin despite their retarded development suggesting a lesion in auxin homeostasis.

II.2.2 - Results

II.2.2.1 - *DASH* encodes a DOF transcription factor preferentially expressed in endosperm

In a previous study (Verdier et al., 2008) a DOF TF gene (Medtr2g014060.1, *DASH*) was identified due to its seed expression restricted to the endosperm. The preferential seed expression was confirmed using the comprehensive dataset available at the Medicago Gene Expression Atlas webserver (MtGEA, <http://mtgea.noble.org>) using the probeset Mtr.21255.1.S1_at corresponding to the gene (Benedito et al., 2008). The phylogenetic relationship between DOF genes from Medicago, Arabidopsis (Yanagisawa, 2002) and soybean (Wang et al., 2007) was established by the neighbor-Joining method. Pea (PsDof) and the cereals DOF proteins (ZmBPF, BPBF and WPBF respectively from maize, barley and wheat) were included in the analysis. The corresponding tree divided into eight clusters (A to H; annexe II-1), which partially match with the classification of Arabidopsis DOF proteins (Yanagisawa, 2002). *DASH* protein, which belongs to group E, clustered with several soybean and pea DOF TFs and with a single gene from Arabidopsis, AtDof1.4 (At1g28310). Interestingly, none of the members of this group currently have a function ascribed to them, and their tissue-specific expressions are not known to be seed preferential, as for the most closely related *M. truncatula* gene, Medtr2g014170.1, which is mainly detected in flowers (Noguero et al., 2013).

The seed developmental stages associated with *DASH* expression was characterized. *DASH* relative expression was determined in seeds dissected from pods harvested at 4, 6, 8, 10, 12 and 14 days after pollination (dap) from A17 genotype by quantitative Real-Time-PCR (qRT-PCR) (Fig. II-1A). *DASH* mRNA was first detected at 6 dap, corresponding to the globular stage in which embryo cell-type differentiation begins. Its expression increased by 8-10dap (late globular-heart stage) to reach a maximum abundance

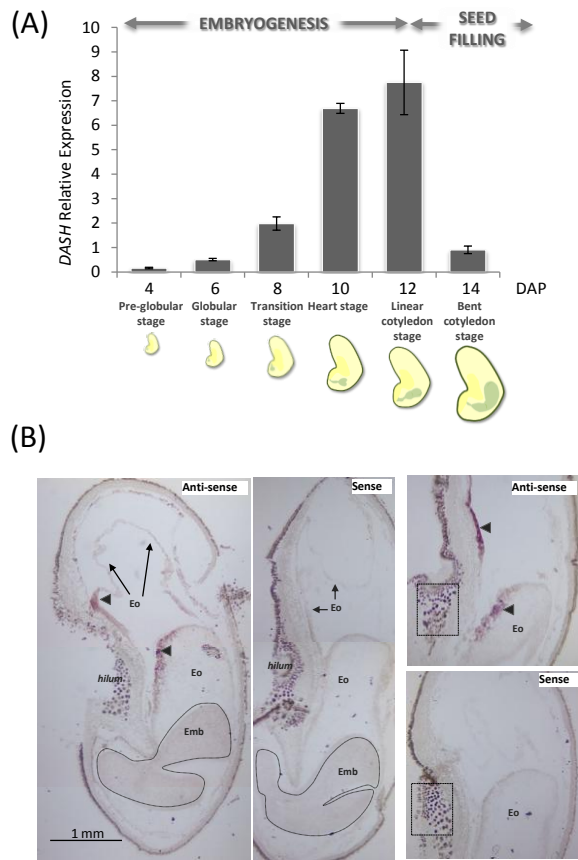


Figure II-1 - *DASH* expression on seed tissue. (A) *DASH* relative expression levels in *M. truncatula* Jemalong A17 genotype through seed development. *DASH* relative transcript abundance was measured by qRT-PCR through the seed developmental on a time scale in days after pollination (DAP). Plants were grown under greenhouse conditions and pods were harvested from 4 DAP to 14 DAP. Expression levels were normalized with 3 constitutive genes (Actin, GAPDH and PDF2, Verdier et al. 2008). Data are average value \pm SD from 3 biological replicates carried out in technical triplicates. (B) In situ hybridization in 12 DAP seed using a *DASH*-specific antisense probe. The signal can be seen as a purple color in the chalazal endosperm (arrowheads). Negative control with the sense probe did not yield to any signal in the endosperm. Dotted squares show background in hilum region, found with both sense and anti-sense probe. Eo = endosperm, Emb = embryo.

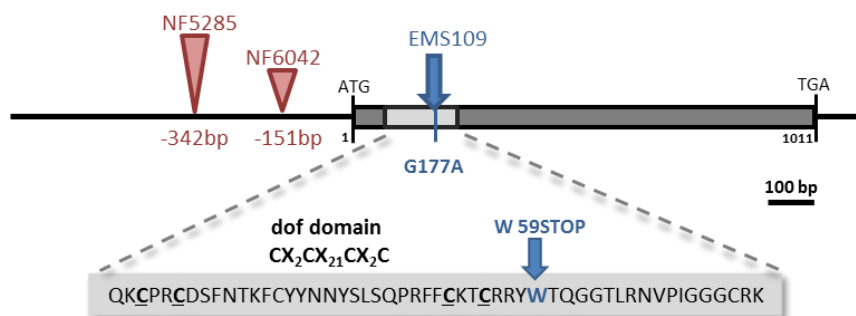


Figure II-2 - Position of the EMS mutation and of the *Tnt1* insertions within the *DASH* sequence. In the EMS109 mutant (in blue, A17 genotype) the mutation W59STOP is located into the DOF domain (light gray box). The *Tnt1* insertions (in red, R108 genotype) are located at -342bp and -151bp from the ATG, for NF5285 and NF6042 respectively.

at 12 dap (linear cotyledon stage), a key transition point between embryogenesis and seed filling. At this stage, after extensive cell divisions, seed storage proteins start to accumulate along with embryo cell expansion (Gallardo et al., 2007). The same expression profile was observed for *DASH* in the R108 genotype, with a peak of expression at 12 dap (data not shown), suggesting a role for this TF during early stages of seed development, and particularly during late embryogenesis.

Further exploration of *DASH* expression within the endosperm was carried out by *in situ* hybridization on 12 dap A17 seeds. The endosperm appeared split in two parts: one attached to the seed coat and another surrounding the embryo, as previously observed for early developing *M. truncatula* seeds (D'Erfurth et al., 2012) whose endosperm is delicate and hence difficult to keep intact during steps of tissue preparation for *in situ* hybridization. Nevertheless, our *in situ* experiments indicated that *DASH* expression was restricted to a specific zone of the endosperm corresponding to the chalazal region of endosperm (CZE) close to the hilum (point of attachment of the seed to the pod) (Fig. II-1B).

II.2.2.2 - *dash* loss-of-function alleles have a seed-lethal or near-lethal phenotype

A *M. truncatula* EMS collection derived from Jemalong A17 genotype (Le Signor et al., 2009) was screened for Targeting Induced Local Lesions IN Genomes (TILLING) (McCallum et al., 2000). One of the mutations (G177A, Fig. II-2) generates a stop codon (W59STOP) located just after the loop within the DOF domain. This may affect protein function because of the dramatic change in protein sequence, including truncation of the DOF domain, resulting in a predicted peptide length of only 59 amino acids instead of 336 for the full length protein sequence (Fig. II-2). This mutant was named EMS109. In addition, two insertional-mutant lines were identified from the *Tnt-1* insertion mutant population generated in the R108 genotype (d'Erfurth et al., 2003) and provided by The Samuel Robert Noble Foundation (Tadege et al., 2008). These insertions were located in the putative promoter region, at -342bp (mutant NF5285) and at -151bp (NF6042) upstream of the ATG start codon (Fig. II-2). The EMS and Tnt1 mutants were back-crossed twice with the corresponding wild-type lines before analysis.

The phenotypes of the different mutant lines compared to their respective WT lines, were studied. While NF6042 and EMS109 exhibited seed-lethal phenotypes, NF5285 displayed a small seed phenotype (Table II-1, Fig. II-3 and annexe II-2). The mutant line NF5285 was found in the homozygous state and produced normal plants with a reduction of 16% of the mature seed weight. Plants heterozygous for the NF6042 insertion were analysed by PCR-based genotyping but no homozygous plant could be recovered from the progeny. Among 124 progenies from self-fertilized heterozygous, 79 were heterozygous and 45 wild-type (annexe II-2II). The observed 2:1 segregation ratio and the failure to obtain homozygous plants suggested that NF6042 was embryo-lethal. To investigate this further, young developing pods of self-fertilized NF6042 heterozygous plants were dissected and observed under the binocular microscope. Small seeds as well as dry seeds were observed within pods in a proportion close to 3:1 (annexe II-2II),

Mutant	Genotype	Seed weight	Seed weight reduction/ WT	Pod weight	Number of seeds per pod
EMS109	<i>dash / dash</i> ^a	1.43 ± 0.042 *	67%	35.5 ± 1.54 *	6 ± 0.23 *
	<i>dash / +</i>	3.99 ± 0.057 *	9%	61.1 ± 1.84 *	6 ± 0.28 *
	<i>+ / +</i>	4.37 ± 0.072		73.8 ± 2.07	8 ± 0.24
NF5285	<i>dash / dash</i>	3.87 ± 0.187 *	16%	127.2 ± 5.12 *	6 ± 0.3 *
	<i>+ / +</i>	4.61 ± 0.103		169 ± 8.25	7 ± 0.3
NF6042	<i>dash / +</i> ^b	3.53 ± 0.07 *	10%	71.58 ± 5.1	5 ± 0.45 *
	<i>+ / +</i>	3.94 ± 0.05		76.8 ± 3.9	8 ± 0.24

(a) seed production in the late stage of the plant life cycle

(b) homozygous has a seed-lethal phenotype

* significant value relative to wild-type ($p < 0.05$, student t-test)

Table II-1 - Description of EMS and TNT mutations obtained for the Jemalong A17 and R108 genotype and associated phenotype.

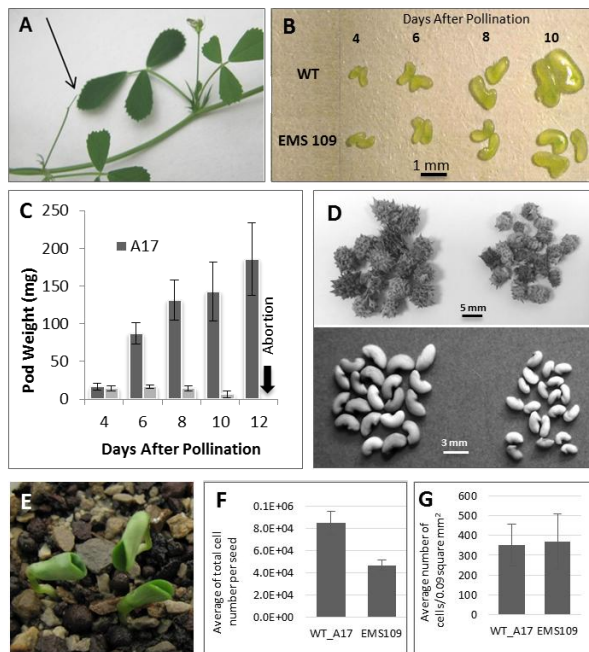


Figure II-3 - Phenotype of the homozygous EMS109 mutant. (A) Pod abortion during the first months of the plant life cycle. (B) Seed development from 4 to 10 dap for wild-type (top) and EMS109 mutant (bottom). (C) Kinetics of pod weight from 4 dap to 12dap, for wild-type and EMS109. (D) Comparison of pod and seed sizes from WT (left) and EMS 109 mutant (right) harvested at maturity at the end of the plant life cycle. (E) Abnormal mutant seedlings showing cup-shaped cotyledons. (F) Average of total cell number per seed from WT and EMS 109 mutant at the mature stage. (G) Average number of cells/ 0.9 square mm from WT and EMS 109 mutant at the mature stage.

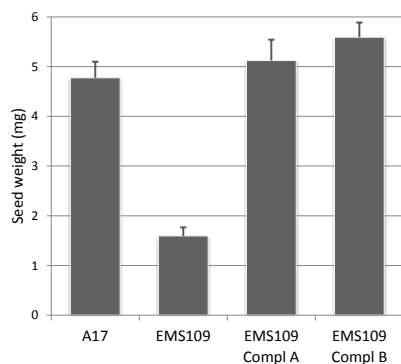


Figure II-4 - Complementation of EMS109 mutant. Mature seed weight respectively measure for A17, EMS109 mutant and two complemented lines: EMS109 compl A and EMS 109 compl B.

consistent with Mendelian segregation of a single recessive allele and confirming homozygous embryo-lethality. Mature pods from selfed heterozygotes showed reduced seed number (Table II-1), consistent with the segregation of an embryo-lethal allele.

Same genetic analyses were performed for the EMS109 (W59STOP) segregating population and lead to similar results (annexe II-2). However, out of 100 progenies from self-fertilized heterozygous plants, one homozygous plant could be obtained, grown and compared to WT. We observed that the EMS109 mutation in the homozygous state resulted in pod abortion during the first months of the plant life cycle: flowers appeared normally but after 3-4 days, flowers and/or small developing pods aborted, dried out and detached (Fig. II-3A). However, after 4-5 weeks with 100% pod abortion, some pods began to develop up to 8-10 days before drying and detaching. These pods were much smaller than the WT pods at the same stage and contained smaller seeds, as shown in Fig. II-3B and C. Interestingly, in the late stage of the plant life cycle (that is after two or three months of flowering), EMS109 homozygous plants were able to produce pods which were smaller than WT, and which contained viable seeds, the weight of which was greatly reduced compared to WT (Table II-1 and Fig. II-3D). The number of seeds per pod appeared also greatly reduced compared to WT plants (Table II-1). These small seeds give rise to abnormal cup-shaped cotyledons phenotype in about 40% of cases (Fig. II-3E). Eighty % to 90% of these abnormal seedlings do not give a viable plant.

II.2.2.3 - Insertion of *DASH* gene in EMS109 complements mutant phenotype

The different *dash* mutations displayed different severities from decrease of seed size (NF5285) to seed-lethal phenotype (EMS109 and NF6042). In order to link these phenotypes, complementation of the *dash* mutation was carried out using the EMS109 mutant line. A construct containing 1150 bp of the native promoter sequence fused with the open reading frame of *DASH* gene was introduced into the pMDC123 vector (Curtis and Grossniklaus, 2003) and transfected into the EMS109 mutant line at the homozygous state using *Agrobacterium tumefaciens*. Among transformant lines containing the *pDASH::DASH* construct selected on the presence of the construct and the selection marker, two lines displayed complementation of the seed lethal phenotype and a restoration of the normal seed weight (Fig. II-4). These complemented lines did not show any vegetative or reproductive phenotype in the next generation such as abnormal cup-shaped cotyledons observed in EMS109 mutants.

II.2.2.4 - *dash* mutants are impaired in embryogenesis

The lethal-seed phenotype prompted us to study in more detail early embryo development of the *dash* mutants. Seeds were dissected from developing pods, cleared and observed using differential interference contrast (DIC) microscopy. A developmental series of WT (A-D) and EMS109 mutant (E-L) cleared seeds from 6 to 12 dap is presented in Fig. II-5. Figures II-5E to II-5H show mutant seeds harvested

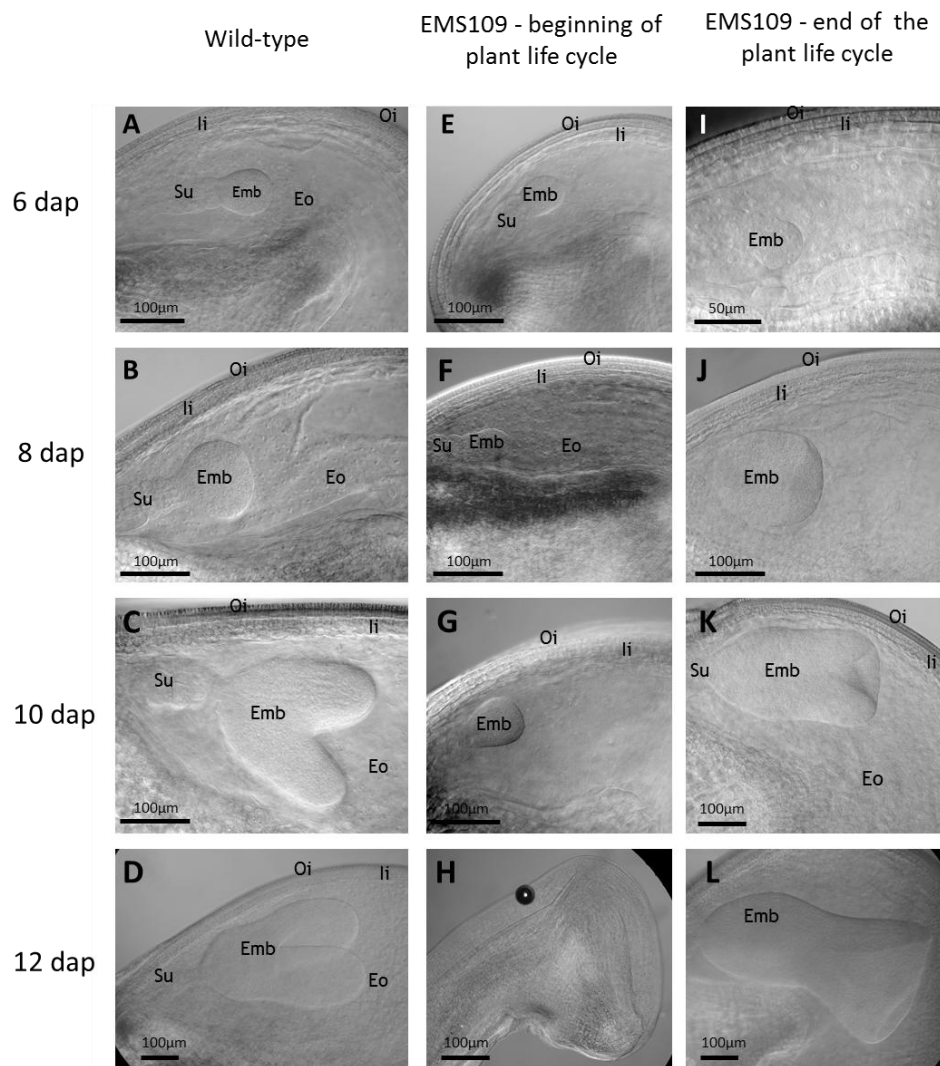


Figure II-5 - Phenotypical characterization of the EMS109 seed lethal phenotype. Cleared pictures of embryo and endosperm through seed development. Wild-type seeds: left pictures (A to D). EMS109 seeds at the beginning of plant life cycle (E to H) and EMS seeds eluded to abortion at the end of the plant life cycle: right pictures (I to L). Globular stage corresponding to 6 dap (A, E, I). Transition stage corresponding to 8 dap (B, F, J). Heart stage corresponding to 10 dap (C, G, K). Linear cotyledon stage corresponding to 12 dap (D, L). Small aborted seed found in EMS 109 pods at 10 dap (H). Emb, embryo; Eo, endosperm; li, internal integument; Oi, outer integument; Su, suspensor. Scale bar indicated on pictures.

from five week-old plants (before pod abortion) while figures II-5I to II-5L show seeds harvested on plants at the end of their life cycle that are able to produce small mature seeds. For seeds harvested from five week-old plants, no change was observed in embryo development at 6 dap (Fig. II-5A and II-5E), whereas embryo development appears delayed at 8 dap (Fig. II-5B and II-5F), coinciding with the beginning of *DASH* gene expression (Fig. II-1). At 10 dap, WT embryos reach heart stage whereas EMS109 embryos were still at a globular stage, attesting embryo growth arrest (Fig. II-5C and II-5G). Cellularized endosperm appears around the embryo between 6 and 10 dap but gradually degenerates, leading to small seeds, tissue collapse and abortion (Fig. II-5H and annexe II-5). Small seeds from developing heterozygous NF6042 pods were also cleared and compared to WT seeds (annexe II-3). Similarly to EMS109, they appeared delayed in embryo development from 6 dap resulting in embryo growth arrest at the globular stage. The EMS109 seeds that escape abortion by the end of the plant life cycle differ from WT seeds by their small size, as previously shown in Fig. II-3D, and the fused cotyledons of the embryo (Fig. II-5I to II-5L). By counting the number of embryo cells per square millimeter in the EMS109 mature seeds that escape abortion, we did not observe any variation compared to the WT, but the total number of cells per seed determined using a hemocytometer after enzymatic digestion was significantly reduced in the EMS109 embryos. These results suggest that cell divisions decreased in the mutant embryos, whereas cell elongation was not affected (Fig. II-3 and II-3G).

II.2.2.5 - A transcriptome comparison of WT and EMS109 mutant pods reveals major transcriptional reprogramming in *dash* mutants

To get an insight into the processes affected in developing *dash* seeds, we used the *Medicago truncatula* Affymetrix GeneChip (Benedito et al., 2008) for hybridization of RNA extracted from the WT and EMS109 mutant pods at 8 dap (about two days before pod abortion). At this stage, both *dash* and WT embryos are in globular stages with *dash* showing a small delay in development (Fig. II-3). Following log transformation of the data and multiple t-test correction, a probe set list of 7453 genes differentially expressed between the *dash* mutant and the WT were selected: 3767 were down-regulated and 3686 up-regulated. The exhaustive list of these genes is presented in supplemental table 1. Schematic PAGEMAN representation of under- and over-represented BIN classes (*i.e.* functional classes (Thimm et al., 2004)) pointed out major effects on most of metabolic pathways (Table II-2). Besides the global effect on the expression of different classes of genes summarized in Table II-2, down- or up-regulation of some of the most affected genes were validated using qRT-PCR (Fig. II-6A and see below). Additional transcriptomic data concerning the three tissues (endosperm, embryo and seed coat) of 12 dap WT seeds were obtained using a Nimblegen array (unpublished data) and are presented when needed to highlight tissue-specific patterns of the transcriptional changes observed in *dash*.

Functional Class		Score	number of genes	Leaf	Petiole	Veg Bud	Stem	Flower	Pod (5-21dap)	Seed_10 dap	Seed_12dap	Seed_16dap	Seed_20dap	Seed_24dap	Seed_36dap	Seed_Coat (16-24 dap)	Root
Down regulated	protein synthesis	-6,8	159	-51	-42	50	-12	-28	33	57	32	15	-1	-20	20	-60	8
	hormone metabolism.auxin.induced-regulated-responsive-activated	-6,0	60	35	33	-10	4	78	10	-17	-21	-24	-25	-23	-23	-11	-8
	hormone metabolism.auxin	-5,8	68	26	29	-7	5	74	12	-15	-19	-21	-23	-22	-23	-9	-7
	cell	-5,4	178	-29	-16	21	5	-14	16	30	18	9	-4	-11	-17	-9	1
	cell.cycle	-5,2	45	-46	-25	39	1	-21	28	55	32	12	0	-14	-28	-31	-1
	DNA.synthesis/chromatin structure	-5,1	89	-37	-31	44	-18	-1	17	62	33	0	-18	-11	-2	-23	-14
	cell.organisation	-3,4	107	-23	-11	12	11	-12	13	21	15	8	-5	-11	-20	0	4
	protein.folding	-3,4	36	-17	-17	56	-3	-24	29	38	17	-3	-19	-29	17	-41	-2
	nucleotide metabolism.salvage	-3,1	10	-36	-22	60	26	-19	56	53	27	-1	-19	-35	-18	-45	-27
	protein.degradation.subtilases	-2,7	24	-15	12	11	29	11	24	33	3	-5	-15	-22	-35	-10	-20
	misc.GDSL-motif lipase	-2,6	29	2	13	47	12	28	12	-1	-15	-17	-19	-13	-13	-17	-19
	cell wall	-2,5	119	-11	22	8	30	21	21	2	-6	-6	-15	-20	-30	-11	-4
	lipid metabolism.Phospholipid synthesis	-2,3	20	74	11	30	-4	19	-1	14	-3	-17	-20	-31	-42	-5	-25
	signalling.receptor kinases.LRR III	-2,2	20	-20	-3	38	36	-4	40	18	8	-6	-19	-25	-42	-29	8
	protein.targeting.nucleus	-2,1	11	-52	-8	40	39	-41	41	59	35	-15	-31	-30	-9	-40	11
	misc.beta 1,3 glucan hydrolases	-2,1	32	-25	-18	21	11	-19	17	28	18	13	-4	-10	-36	14	-9
	misc.plastocyanin-like	-2,1	7	-33	-20	20	13	0	26	38	22	9	2	-4	-31	-27	-15
	signalling.G-proteins	-2,1	54	-33	-17	22	10	-11	20	30	13	4	-6	-13	1	-29	10
	RNA.transcription	-2,0	17	-59	-40	32	-25	-21	-3	51	27	-5	-10	-7	94	-27	-7
Up regulated	hormone metabolism.cytokinin	2,0	8	-6	-18	-34	-13	59	-24	11	3	16	-10	-5	-2	53	-30
	development	2,2	165	-7	-2	-5	6	6	-3	-4	-4	-8	-8	-3	11	8	15
	lipid metabolism.FA synthesis and FA elongation.acyl coa ligase	2,2	33	-6	-5	28	3	-15	4	6	17	28	2	-17	-20	-16	-7
	hormone metabolism.ethylene.induced-regulated-responsive-activated	2,3	11	-21	-13	-30	-16	40	-19	0	-8	-8	-2	-4	7	33	41
	amino acid metabolism.synthesis.central AA metabolism	2,4	6	70	20	9	-9	-17	-23	-16	-10	-6	2	5	-24	5	-5
	hormone metabolism.abscisic acid	2,6	19	27	6	-27	-18	4	-27	-8	-12	-3	-3	2	20	44	-5
	misc.cytochrome P450	2,7	50	19	6	-14	-12	32	-12	-10	-11	-13	0	1	7	-1	8
	amino acid metabolism.degradation	2,7	26	-23	-14	-19	-20	19	-8	-18	-15	2	18	22	29	29	0
	hormone metabolism.ethylene.synthesis.degradation	2,8	30	32	25	-12	-14	15	0	-9	-11	-8	-7	-6	3	7	-15
	protein.degradation.autophagy	2,9	13	-3	-11	-51	-39	37	-36	-28	-16	-19	2	17	31	71	45
	minor CHO metabolism	3,0	26	-27	-22	-24	-19	14	-4	-5	3	5	-4	-1	32	42	10
	protein.posttranslational modification	3,1	104	-2	-4	-12	-6	8	-12	-3	-2	-3	-3	-1	-1	27	14
	transport	3,5	286	3	0	-9	3	24	5	-2	-3	-8	-10	-12	-6	2	13
	hormone metabolism.ethylene	3,6	50	17	13	-8	-10	21	-1	-9	-11	-11	-7	-8	3	8	5
	not assigned	3,7	4380	-6	-5	3	-1	0	4	5	0	-4	-5	-7	9	-1	8
	transport.amino acids	3,9	41	17	16	-24	1	41	-5	-21	-25	-24	-21	-18	16	0	47
	protein.degradation	6,5	32	35	-3	-13	-16	15	-7	3	0	-7	-5	-7	11	8	-13
	protein.degradation.ubiquitin	7,3	160	-15	-10	-31	-14	1	-26	-11	-11	-6	3	11	47	49	12

Table II-2 - PageMan display of coordinated changes of gene categories deregulated 8 dap in pods of *dash* mutant compared to WT. Log₂ ratio of mutant data versus WT data from three repetitions of Affymetrix GeneChip were subjected to a bin-wise Wilcoxon test followed by a Benjamini-Hochberg pvalue correction for multiple testing. Score of the wilcoxon test is presented as well as number of deregulated genes in each class, only classes with Benjamini-Hochberg pvalues > 0.05 were selected. In the display, the overrepresented MapMan functional categories are given by collapsing nonsignificant categories. Mean expression data from Medicago geneatlas (<http://mtgea.noble.org/v3/experiments>) (data from Benedito et al., 2008 and seed coat data from Pang et al., 2008) were standardized within the two batches and then averaged over the deregulated class. The color scale ranged from red (max expression) to green (minimum expression) in the tissue.

II.2.2.6 - Down-regulated genes in *dash*

Protein synthesis (mainly ribosomal proteins), auxin pathways and auxin response genes, cell cycle (cyclin-dependent protein kinase regulator) and chromatin structure (nucleosome assembly, DNA binding or DNA replication factor and histone synthesis) were the most down-regulated functional classes in *dash* (score <-5, Table II-2). Expression profiling data for developing seeds and other plant tissues were retrieved from the MtGEA webserver (WT backgrounds) for the deregulated genes in each class. Then, the mean expression value per class was calculated and presented in Table II-2. Interestingly, this revealed that the down-regulated probes in *dash* were mainly strongly expressed at early stages of seed development in WT seeds (Table II-2). The remaining down-regulated classes included a wide variety of cellular functions needed for cellular development, among which are the lipid metabolism class with phospholipid synthesis involved in cell membrane modifications and the cell wall class with many probes of pectate lyases and polygalacturonases. Among them the probe Mtr.26579.1.S1_s_at, corresponding to the seed specific polygalacturonase Medtr7g088550.1, was confirmed by qRT-PCR to be down-regulated in 8 dap *dash* seeds and showed a seed specific expression in the endosperm (annexe II-5).

Besides the down regulation of specific metabolic pathway/classes, we searched for the most strongly down-regulated genes in *dash* (\log_2 expression ratio between *dash* and WT <3.5, expression reduced by at least 11 fold). A smaller number of genes (13) was identified, of which eight displayed an expression pattern correlated to *DASH* expression (*i.e.* seed specific with an expression peaking between 10 and 12 dap, Fig. II-6A). Moreover, our seed tissue array hybridization confirmed that all of them were preferentially expressed in endosperm or in both endosperm and seed coat (Fig. II-6C). The five most deregulated probes (expression reduced by 40 to 75 fold) were legume-specific according to the *Medicago* gene atlas (Benedito et al., 2008). Two of them (Mtr.1254.1.S1_at, Mtr.21576.1.S1_at) correspond to short cysteine-rich proteins resembling defensins (defensin-like proteins, (Tesfaye et al., 2013)), which are antimicrobial polypeptides often associated with boundary cell layers. Down-regulation of these two probes was confirmed in *dash* seed tissues at 8 dap (annexe II-6 or Fig II-6D) and expression of the most deregulated one (Mtr.21576.1.S1_at) was shown by *in-situ* hybridization to be restricted to the chalazal endosperm, with a pattern identical to that of *DASH* (compare Fig II-6B and II-1B), suggesting this defensin could be a target gene of *DASH*. The three remaining probes (Mtr.19455.1.S1_at, Mtr.2563.1.S1_at and Mtr.50101.1.S1_at whose down regulation in *dash* seed was validated by qRT-PCR in annexe II-6) lack clear homologs by Blast, but possess signal peptides and several cysteine residues suggesting they may also have a defensin-like function. A further three probes with expression patterns similar to *DASH* and down-regulated by 15- to 35-fold in *dash* compared to WT were identified as non-specific lipid-transfer protein ns-LPT_like (Mtr.22353.1.S1_at), germin-like (Mtr.39975.1.S1_at) and desiccation-related protein (DRP, Mtr.42115.1.S1_at) (Fig. II-6A).

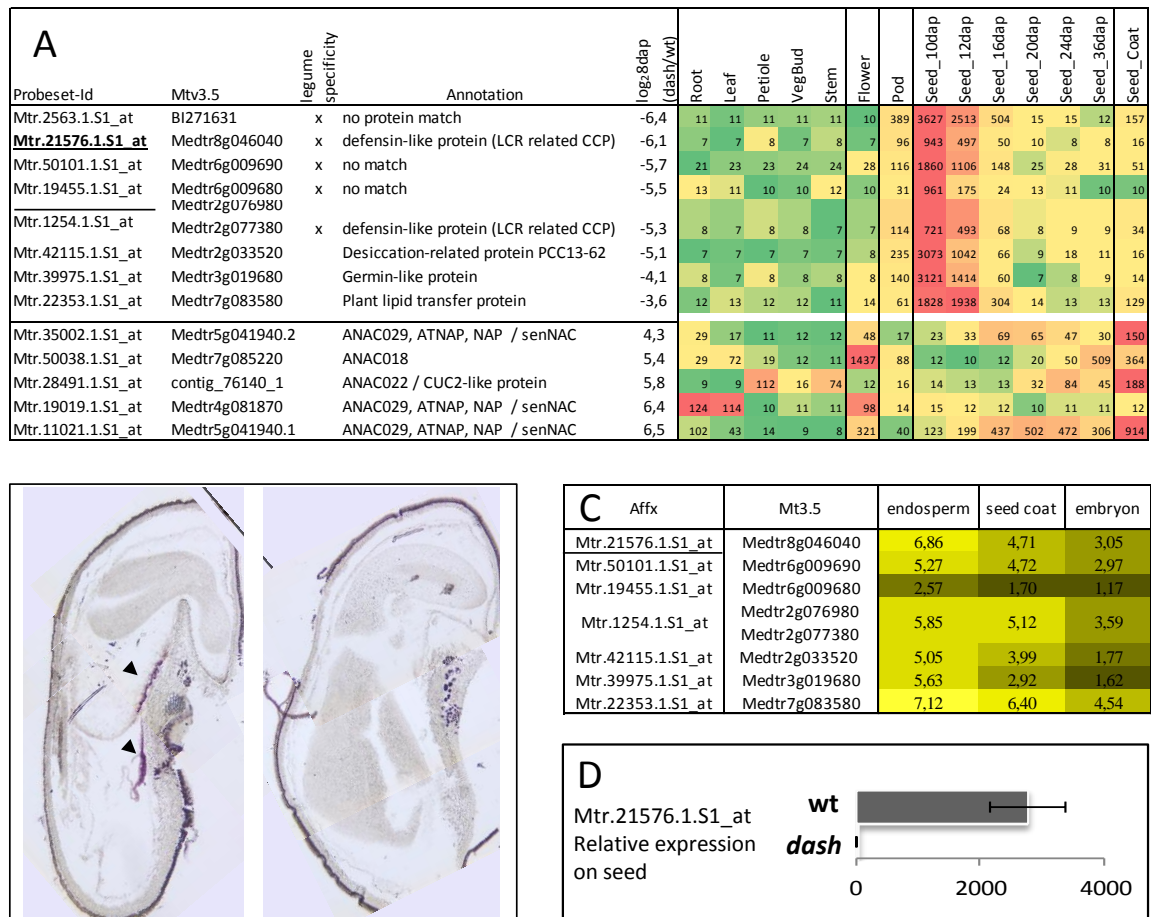


Figure II-6 - The most deregulated genes in 8 dap pods of *dash* mutants vs wt. (A) List of the 8 most down-regulated genes (log₂ ratio of 8 dap pod expression of mutant vs wt) showing a seed specificity pattern close to dash one and of the five transcription factor belonging to NAC family most up-regulated in dash. Genome identification V3.5, gene annotation and indication of legume specificity were added. Data from Medicago geneatlas (Benedito et al., 2008) were presented (plant tissues and seed development as well as seed coat data from Pang et al. (Pang et al., 2008)). (B) In-situ hybridization of defensin probe Mtr.21576.1.S1_at showing expression in the chalazal region of endosperm (arrowheads). (C) Tissue specificity for down regulated probes (unpublished data). (D) qRT-PCR relative expression of Mtr.21576.1.S1_at in 8 dap seeds of dash and wt.

II.2.2.7 - Up-regulated genes in *dash*

Protein degradation (mainly E3 ubiquitin ligases) and transport classes (mainly amino acid transporters orthologous to amino acid permeases, cationic amino acid transporters and proline transporter) were the most up-regulated functional classes in *dash*. These probes were mainly expressed in WT flower and at the end of seed development, and are thus expressed ectopically in young *dash* seeds (Fig. II-6A). Among the most up-regulated genes, we found five NAC [for NAM (no apical meristem), ATAF, CUC (cup-shaped cotyledon)]-domain TFs, three of which have similarities with the AtNAP gene (NAC-like, activated by APETALA3/PISTILLATA). These genes were up-regulated in *dash*: 25-fold for Mtr.35002.1.S1_at, 50-fold for Mtr.50038.1.S1_at and Mtr.28491.1.S1_at and 70-fold for Mtr.11021.1.S1_at and Mtr.19019.1.S1_at (supplemental table 1 and Fig. II-6A). These TFs are normally expressed in WT flowers and/or during seed filling (16-36 dap) but weakly expressed in early stages of seed development (Fig. II-6A), suggesting they could be repressed by *DASH*.

II.2.2.8 - A striking deregulation of hormone metabolism in *dash* mutant

Genes related to the functional class “hormone” and particularly those related to auxin action, were among the most deregulated in *dash* mutants. Interestingly, most of the auxin-related probes were down-regulated in *dash* (Table II-2 and annexe II-6). Among them were orthologs of the auxin efflux carriers AtPIN1 (MtPIN10 and MtPIN4), AtPIN7 (MtPIN1 and MtPIN3) and AtPIN8 (MtPIN8). Two of them (MtPIN3 and MtPIN4) are mainly expressed in the embryo tissue (annexe II-6). Significantly, 65% of the down-regulated genes in the auxin class encode auxin-induced proteins (Table II-2). In the MtGEA, these auxin-responsive genes appear preferentially expressed in pods compared to seeds, suggesting they may also play an important function within the pod. Interestingly, among the few auxin-responsive genes up-regulated in *dash* is *IAA18-LIKE* which alters PIN1 expression in Arabidopsis. Gain of function mutation in this gene alters *AtPIN1* expression in the apical domain of the embryo, indicating that IAA18 disrupts auxin transport (Ploense et al., 2009).

In contrast to auxin-related genes, genes involved in the metabolism of abscisic acid (ABA), ethylene and cytokinins were mainly up-regulated (Table II-2). Among the up-regulated probes related to ABA were ABA1, ABA3 and AAO3-related sequences and two sequences of ABA response genes (*AREB1*, *ATBZIP12*). Among the up-regulated probes in *dash* related to ethylene were several sequences orthologous to ACC oxydase and ACC synthase, suggesting an activation of ethylene synthesis. A histidine kinase sequence related to *EIN1* (*ETHYLENE INSENSITIVE 1*; Mtr.43466.1.S1_at) and many members of the AP2/ERF transcription factor family (ERF9, ERF1, ERF6, ERF11) were also up-regulated (up-regulation of probe Mtr.45730.1.S1_at was confirmed in *dash* seeds, annexe II-6). All ethylene-related probes were normally mainly expressed in WT leaves, flowers and/or during late stages of seed development. Regarding the cytokinin-related probes, cytokinin oxidase/deshydrogenase genes, such as putative orthologs of *ATCKX6* and *ATCKX5*, were up-regulated in the *dash* mutant, suggesting increased cytokinin degradation. Three

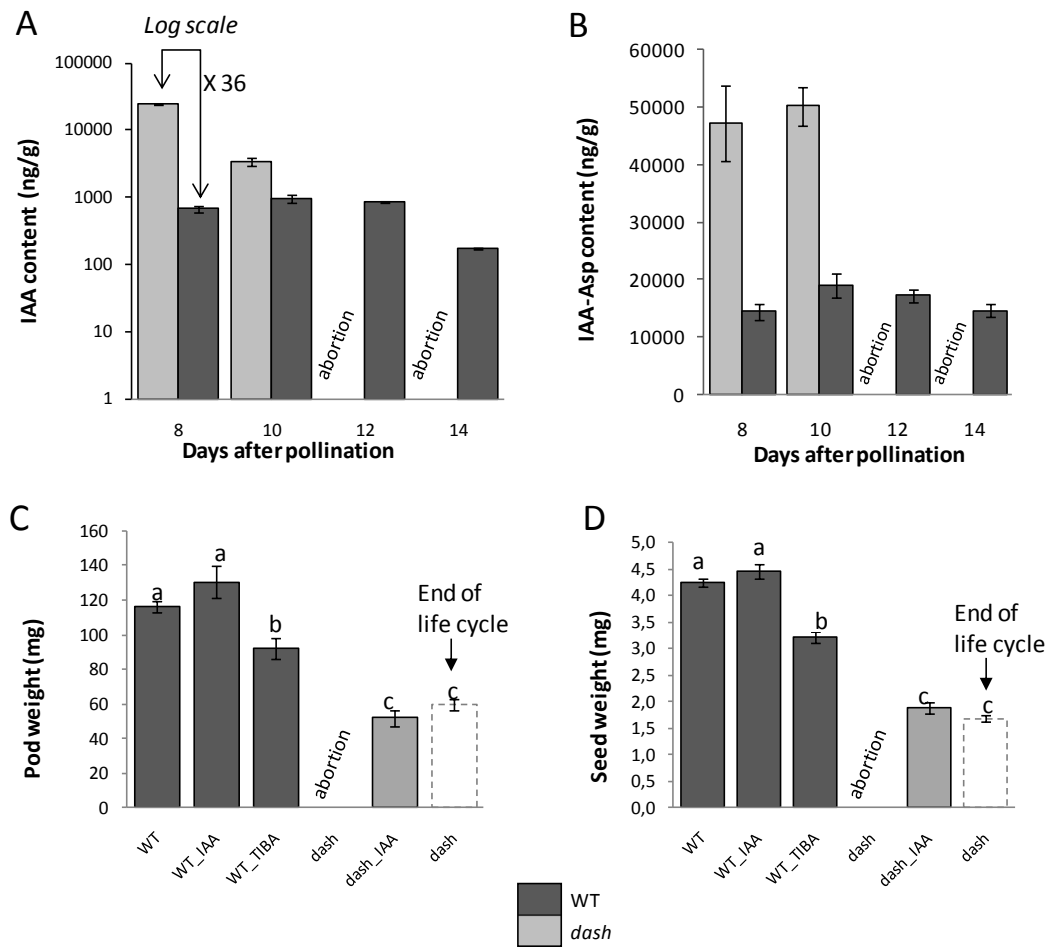


Figure II-7 - Hormone dosage and treatment effect on pods of *dash* vs wt. (A and B) Hormone dosage in pods of 8 dap of EMS109 and WT. (C and D) Effect of IAA and TIBA (inhibitor of auxin efflux) treatments on 6 dap pod on pod and seed weights of EMS109 and WT. Concentration applied for IAA and TIBA: 100mg.L-1. A minimum of 20 pods and containing seeds were weighted.

cytokinin-O-glucosyltransferase genes (Mtr.42515.1.S1_at, Mtr.40385.1.S1_at, Mtr.46619.1.S1_x_at) involved in cytokinin storage and homeostasis (Hou et al., 2004) and two probes orthologs of AHK3 (Mtr.32270.1.S1_at, Mtr.9978.1.S1_at), involved in cytokinin signaling, were also up-regulated. Despite a relatively high number of transcription factors deregulated in *dash* ($328/4172 = 8\%$ versus around 5% in the genome), any particular TF class was significantly deregulated and disparity was observed between different TF families. When applying a cutoff of 4-fold change between WT and mutant, we observed that NAC, AP2-EREBP, BZIP, BHLH classes were up-regulated. Applying this cutoff, only the Aux/IAA family class was down-regulated confirming general down-regulation of auxin signaling in *dash*.

II.2.2.9 - A link between DASH and auxin signaling

In order to better understand the link between the *dash* mutation and the deregulation of auxin responsive genes observed in the mutant transcriptome, auxin content was measured in EMS109 mutant and WT pods using liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). At 8 dap, the stage used for transcriptomic analysis, IAA content was about 36 times higher in *dash* pods compared to WT [figure II-7A]. Higher IAA content in mutant pods was still observed at 10 dap, just before mutant pod abortion which occurred around 12 DAP. In WT pods, IAA content started to decrease at the onset of seed filling (14 DAP). Interestingly, we observed a higher accumulation of the auxin conjugate IAA-Asp in 8 and 10 dap mutant pods compared to WT, which corresponds to an inactive form of auxin (Fig. II-7B). Same observation was made for the NF5285 mutant pods, which displayed 11-fold higher content of auxin than WT pods (data not shown).

In view of the abnormal accumulation of auxin in mutant pods during early embryogenesis, we tested the effects of an inhibitor of auxin transport on seed development and growth. TIBA (2,3,5-triiodobenzoic acid), an inhibitor of auxin efflux, was used at 100 mg.L^{-1} to treat WT pods at 6 dap corresponding to the onset of *DASH* expression. We observed a significant decrease in mature pod and seed weights in response to TIBA (Fig. II-7C and D), suggesting that auxin is required for normal seed development and that disturbed auxin transport could be a partial explanation of impaired EMS109 seed development. No change in the number of seeds per pod was observed (data not shown). In order to get an insight into the effect on auxin transport, we treated *dash* and WT immature pods at 6 dap with natural auxin IAA. As described by Delbarre et al (1996) (Delbarre et al., 1996), natural forms of auxins such as IAA are strictly imported into the cells by the LAX influx transporters and exported outside the cells by the PIN efflux transporters. IAA treatment did not cause significant changes in pod and seed weight in WT, but caused a recovery of pod weight as well as seed weight for the EMS109 *dash* mutant similar to that obtained at the end of plant life cycle. In conclusion, TIBA treatments mimics the *dash* phenotype on WT, and exogenous IAA applied to pods at 6 dap is able to enhance pod and seed set, suggesting that in the mutant pods auxin efflux transport (*i.e.* involving the PIN transporters) might be impaired.

II.2.3 - Discussion

The *Medicago truncatula* gene *DASH*, which belongs to the DOF family, is specifically expressed in the endosperm during embryogenesis (Fig. II-1). To assess more precisely the relationship between *DASH* and other DOF transcription factors from several species, the full-length protein sequences were compared and a phylogenetic tree was constructed, which identified several related groups (annexe II-1). *DASH* was grouped in a clade with several sequences from *A. thaliana*, soybean or pea, but none of these genes have had a function ascribed to them, and their sites of expression, when known, differ from *DASH*. To investigate the role of this endosperm-specific DOF transcription factor we have screened both the TILLING and Tnt1 populations of *M. truncatula* (Tadege et al., 2008; Le Signor et al., 2009). Two loss of function mutants were identified for this gene, one producing a truncated form of protein (EMS109) and the other carrying a Tnt1 insertion in the promoter (TntT6042). Both mutants displayed altered embryo phenotypes during seed development (Fig. II-5 and annexe II-3), starting with a delay in embryo growth and leading to growth arrest at the globular stage and premature pod abscission. Hence, *DASH* expression in the endosperm plays an important role in controlling embryo growth. This is reminiscent of previous data reporting that endosperm cellularization and embryo development are closely related, and probably link *via* nutrient translocation to determine final seed size (Hehenberger et al., 2012). In contrast to the *mini3* mutant of *Arabidopsis*, which shows precocious endosperm cellularization probably responsible for a reduced seed size (Luo et al., 2005), cellularized endosperm in EMS 109 mutant is present around the embryo between 4-6 dap and 8-10 dap (Fig. II-5E and F) but gradually degenerates, leading to tissue collapse and desiccation, and finally to small seeds (Fig. II-5G and H and annexe II-3).

At the end of screening for EMS population, only one homozygous plant (EMS109) have been rescue from self fertilization of heterozygous plants. Mature plants of EMS109 progressively acquire the ability to develop pods and set seeds, albeit small ones, suggesting a partial physiological reversion of the mutant phenotype (Fig. II-3 and Fig. II-5K and L). We hypothesize here that this reversion involves a modification in auxin transport or perception. Indeed, observations of morphological changes in embryo shape and cup-shaped cotyledons have already been demonstrated as linked to auxin, 2-4D, or auxin inhibitor NPA, and a model involving the role of auxin transport in the control of embryo growth has been put forward (Friml et al., 2003).

To characterize the changes in gene expression occurring in EMS109 mutant seeds, a comparison of the transcriptome of developing seeds from wild-type and EMS109 mutant lines was carried out using the *Medicago truncatula* AffymetrixGeneChip (Benedito et al., 2008). This study has revealed major reprogramming in the mutant seed affecting notably genes involved in hormone metabolism, protein synthesis and degradation, cell cycle and DNA/chromatin remodelling (Table II-2).

II.2.3.1 - *DASH* is expressed in the chalazal endosperm CZE and deregulates a series of genes specifically expressed in the developing endosperm

By comparing the transcriptome of the wild-type and *dash* mutant pods, a class of genes was identified that were specifically expressed in the endosperm with kinetics similar to *DASH* and whose expression was greatly reduced in the mutant (Fig. II-6A and C). These genes encode small cysteine-rich defensin-like (DEFL) peptides and interestingly, one of them (Medtr8g046040) was found by *in situ* hybridization to be expressed in the chalazal endosperm, in a pattern identical to *DASH*, suggesting that this defensin could be a target of *DASH* (Fig. II-6B and Fig. II-1B). DEFL are well known antimicrobial peptides, but some members of this highly diverse family are also implicated in plant development, and in particular in processes such as fertilization, female gametophyte or seed development (Marshall et al., 2011). In maize, the short cysteine-rich peptide MEG1 was shown to be specifically expressed in the endosperm and more precisely in the BETL (Basal Endosperm Transfert Layer, functionally equivalent to the chalazal endosperm CZE in Arabidopsis). *Meg1* regulates maternal nutrient translocation into the seed and seed size, as RNAi lines display smaller seeds. The Medicago genome contains 682 *DEFL* genes, of which 52 are exclusively expressed in developing seeds (Tesfaye et al., 2013). Whether some of these small peptides, and in particular the putative *DASH* target, could have a role in early endosperm development remains to be studied.

The exclusive localization of the *DASH* mRNA in the CZE is of a particular interest. Indeed this endosperm domain has been shown to be a unique subregion, highly enriched for mRNAs specifically expressed during early seed development, suggesting that novel processes occur in this zone (Belmonte et al., 2013). The authors showed that the CZE is enriched in transcripts for genes of ubiquitin-dependent protein catabolism and of enzymes for the biosynthesis of the hormones gibberellic acid, abscisic acid, and cytokinin, consistent with other reports showing that hormone metabolism genes are expressed in the CZE (Miyawaki et al., 2004; Hu et al., 2012). Because of the importance of these hormones for seed development, it was proposed that the CZE may serve as a communication hub that integrates developmental processes within the seed (Belmonte et al., 2013). The misregulation of CZE-specific genes in *dash* suggests that this endosperm sub-region is severely affected, and is also consistent with associated developmental effects on the embryo, which depend on the endosperm at this stage.

In addition to defensin-like peptides, many genes expressed early in seed development and involved in cell division were down-regulated in *dash* (Table II-2), including probes of cyclin-dependent protein kinase regulators (genes orthologous of CYCD, CYCB, CYCA), and cytoskeletal proteins (kinesin and phragmoplast-associated kinesin, tubulin, fimbrin). Cyclin and CDK genes have been shown to be over-represented in proliferating endosperm tissue (Day et al., 2008) as compared to other seed tissue, as syncytial development requires a rapid progression through the cell cycle, suppression of phragmoplast formation

and an coupling of cytokinesis from mitosis. Alteration of cell division in *dash* mutants is confirmed by the lower number of cells observed in cotyledons of mature seeds.

Cell wall deposition was also affected in *dash*, with deregulation of many pectate lyase, polygalacturonase, and xyloglucan endotransglycosylase genes, as well as genes of glycerolipid synthesis needed for membrane formation. These observations are consistent with the impairment of formation of normal cellular structures seen in *dash* endosperm (annexe II-4). Probes involved in DNA synthesis such as histone synthesis, nucleosome assembly, DNA binding or DNA replication factor are mainly down-regulated in *dash*, as are ribosomal protein probes, suggesting reduced anabolic activity in *dash*. Among down-regulated DNA probes, 50% were preferably expressed in early developing seeds (supplemental table 1) and confirmed to be expressed in endosperm in our seed tissue array in agreement with Day et al. (Day et al., 2008) in *Arabidopsis* reporting a significant over representation of chromatin-related genes in endosperm during syncytial development.

II.2.3.2 - Dash acts partly by repressing expression in the endosperm of genes normally active in other seed/flower tissues

Inspection of the genes up-regulated in *dash* mutant pods indicates that DASH, probably in association with other proteins, may also act as a repressor of gene function, similarly to that previously observed for the HvDOF19 transcription factor that mediates the ABA-dependent repression of hydrolase genes in germinating barley aleurone (Moreno-Risueno et al., 2007). The genes affected are in general not expressed in WT endosperm at 10 DAP, but either in pods or flowers, or in some cases in the seed but at the end of the maturation period. Inactivation of *DASH* results in their expression during early seed development. Among the genes most strongly up-regulated, protein degradation, transport, minor CHO metabolism and ethylene and ABA classes were found. Some of these pathways may be interconnected as different E3 ubiquitin ligase gene sub-families are regulated by auxin, gibberellins, ethylene and ABA, and most characterized to date act in regulating plant development and hormone signaling (Mazzucotelli et al., 2008). Their overexpression may contribute to the defective syncytial development of endosperm in *dash*, and the defective seed maturation step.

II.2.3.3 - In the absence of DASH, defective endosperm development elicits a generalized stress response

In the absence of *DASH*, endosperm development seems to be retarded, and from 10 DAP onwards, the endosperm cells visibly degenerate (Fig II-5 and annexe II-3). We hypothesize that, as a consequence, a generalized stress response occurs, effective in the endosperm itself and perhaps also in the interacting tissues, the embryo and the seed coat. The changes in gene expression seen are consistent with this hypothesis: three of the six up-regulated NAC genes are annotated as senescence-related, and their induction in *dash* would be consistent with a role in the degeneration of the endosperm observed. To support this hypothesis, further gene classes up-regulated in *dash* are stress responsive. These include

degradation of amino acids, protein turnover and desiccation-related carbohydrate accumulation (RFO and trehalose) as well as the down-regulated nucleotide metabolism class (Table II-2).

II.2.3.4 - Deregulation of hormone homeostasis in *dash*

a - Auxin efflux and signaling are impaired in *dash* seeds.

The high concentration of IAA measured in *dash* mutant pods (Fig. II-7) together with the fact that the expression of genes involved in auxin biosynthesis/catabolism was not significantly affected in *dash* whereas auxin-induced genes were down-regulated suggests that auxin transport/distribution may be impaired in the mutant. A similar elevated concentration of auxin in seed and fruit was observed upon NPA treatment in tomato fruit (Pattison and Catala, 2012) at an early stage of development. The basipetal polar auxin transport from the fruit to the parent plant is believed to prevent premature abscission (Berry and Bewley, 1991; Serrani et al., 2010). In the *dash* mutant, pod abscission occurs and pod growth and seed set are recovered upon external application of auxin (AIA) directly on immature pod thus possibly restoring an auxin gradient. The existence of auxin gradients during fruit growth, with the highest levels of auxin in the seeds (Gustafson, 1939) is well documented but increasing evidence suggests that after fruit set, precise spatial and temporal synthesis, transport and action of auxin are required for fruit development (Sundberg and Ostergaard, 2009). Auxin distribution is controlled by polar auxin transport mediated by PIN and AUX/LAX proteins, which control cellular auxin efflux and influx respectively (Vanneste and Friml, 2009) and thus the establishment of auxin gradients promoting auxin signaling (Lau and Deng, 2010). The probable orthologs of auxin efflux carriers are mainly down-regulated in *dash* (annexe II-6 and supplemental table 1) and their inactivation is known to result in embryo lethality or severe apical defects depending on the ecotype background (Friml et al., 2003). In *dash*, the polarity of the auxin gradient in the embryo is presumably restored naturally by the end of the plant's life cycle or earlier by external application of IAA. We deduce that *dash* mutants are defective in auxin transport – this leads to auxin accumulation in the immature pod, but due to lack of perception or transport auxin-responsive genes fails to be activated so severely disturbing seed and pod development. The nature of the relationship between *DASH*, expressed in endosperm and deregulation of embryo growth, putatively through auxin-driven development, remains to be clarified.

b - Other hormone signal transduction and metabolism pathways impaired in *dash* seeds

Other hormone metabolism pathways were also affected in *dash*. Absciscic acid (ABA) metabolism-related sequences were mainly up-regulated (annexe II-6). Several sequences related to EFE (Ethylene Forming Enzyme) or ACC synthase were up-regulated in the *dash* mutant suggesting the possibility of an elevation in ethylene in *dash*. Ethylene and ABA are synthesized prior to, and promote fruit set (Vriezen et al., 2008). Many stress-inducible genes are also regulated by endogenous ABA (Yamaguchi-Shinozaki and

Shinozaki, 2006). High levels of ABA and ethylene in the *dash* mutant probably reflect retarded seed development and may contribute to eliciting a general response to stress. Finally, up-regulation of enzymes involved in cytokinin conjugation and degradation in *dash* is consistent with reduced cell division and the severe effect on endosperm development seen in the *dash* mutant.

II.2.4 -Materials and methods

II.2.4.1 - Plant growth conditions and harvest of developing pods

All plants were grown under greenhouse conditions in 1.5 liter pots filled with a mix of attapulgate: clay beads (60:40). Temperature was controlled to be 20°C during the day for both species and above 18°C during the night. Artificial lighting was supplied to reach 16 hours light per day. The plants were not inoculated with *Sinorhizobium sp.* bacteria and nitrogen supply was non-limited: plants were automatically supplied with 3.5N/3.1P/8.6K.

Individual flowers were tagged and seeds collected from 4 to 30 dap according to Gallardo et al. (2007) (Gallardo et al., 2007). Seeds were dissected under a binocular microscope on Petri dishes placed on ice. All seed samples were ground in liquid nitrogen using mortar and pestle. The powder was stored at -80°C until RNA extraction.

II.2.4.2 - Screening for *dash* mutant lines

M. truncatula EMS109 mutant was identified from an EMS population containing 4600 M2 lines from cultivar Jemalong A17 by TILLING screening according to Le Signor *et al.* (2009) (Le Signor et al., 2009). One amplicon of about 1kb was screened to cover the entire coding sequence. Nested PCR was conducted using the following inner primers labelled with IRD-700 and IRD-800 dyes: MtDOF-F2 5'-CCAACCAATAGCAGTAGCAACCG-3' and MtDOF-R2 5'-GCTGCACCTAGAAAATCCAAAGAT-3'. Eight missense mutants and one stop codon mutant were identified and sequenced (Millegen, Toulouse, France). The stop codon mutant was backcrossed twice with the WT lines (Jemalong A17). All progenies BC1, BC2, BC1S1 and BC2S1 (selfing of BC1 or BC2) were genotyped with specific dCaps (derived Cleaved amplified polymorphic sequences) marker designed according to Neff et al. 1998 (Neff et al., 1998).

The *M. truncatula* Tnt1 mutant population containing more than 20,000 lines from cultivar R108 was screened as described in Cheng et al. (2014). Two independent mutant lines containing *Tnt1* insertions located in the promoter region of *DASH* gene were identified by nested PCR using forward primers AGGGTCCCATTCTTTGACTAGT and TGACTAGTGCCACCTCATTGTG and reverse primers TCACTGAGGAGGATTGAACTCTG and TACCATTACCACCAGCACCA. *DASH* gene was sequenced in both mutant lines (NF6042 and NF5285) to confirm and locate the exact position of the *Tnt1* insertions. All progenies were genotyped by PCR using the screening primers and a primer located in the *Tnt1* insert

(CTCCAGACATTTTATTTTACCAAG). Homozygote mutant and sibling wild-type lines were multiplied in greenhouse under conditions described above.

II.2.4.3 - Seed trait phenotyping

The plants analyzed were progenies issued from one to three generations of selfing (S1, S2 or S3) after two back-crosses (BC2S2, BC2S3) of the EMS mutant line with the WT A17 or R108 lines (*M. truncatula*). For each mutation, comparisons were made between the seeds from the WT and mutant plants (between four to seven plants of each genotype were studied) coming from the same progeny in two independent experiments between 2009 and 2012. Pods were harvested at maturity, weighed and then manually threshed. The number of seeds per pod was recorded in both species for a sample of 30 pods and then each seed was weighed individually.

II.2.4.4 - Bioinformatics search for new DOF genes and phylogenetic analysis

M. truncatula sequences were retrieved from the Mt3.5 version of the *M. truncatula* genome. The search for potential DOF genes and the corresponding amino acid sequences in the genome of *Arabidopsis thaliana*, *Glycine max*, *Oryza sativa*, *Zea mays* and *Pisum sativum* was done using "Plant Genome Database" (<http://www.plantgdb.org/>) and NCBI websites. Alignment of conserved regions prior to the distance analysis was also carried out by CLC sequence viewer using default parameters. The alignment algorithm used was the built-in algorithm in CLC sequence viewer and the phylogenetic reconstruction method selected was Neighbor-Joining (CLC sequence viewer 6.2 downloaded from <http://www.clcbio.com/> (CLC bio, Aarhus, Denmark).

II.2.4.5 - In situ hybridization

Freshly harvested *M. truncatula* seeds (at 12 dap) were fixed in 4% paraformaldehyde. A 1125bp probe was defined including all CDS and about 50bp in 3' and 5' untranslated region (UTR) with the forward primer 5'- TGGCACGTCTCTGTTTAGAAAACGC -3' and the reverse primer 5'- GATTAATTCTCAGAGCTATGCTTTC -3'. For the defensin gene Mt8g46040, a probe of 521bp was defined the same way as DASH with primers Forward 5'-GTCTTTGTAAACCTTTCAAAGTAGCC-3' and Reverse 5'-TTGAGATTGTACAAATATCAACCAC-3'. Sense and antisense probes were first amplified using T3 polymerase oligonucleotides with a 5' transcription T3 promoter fusion for template production by PCR, and then synthesized using the T3 polymerase. The method for digoxigenin labeling of RNA probes, tissue preparation and *in situ* hybridization was described by Coen et al. (Coen et al., 1990) with modifications described by Bradley et al. (Bradley et al., 1993).

II.2.4.6 - Cell size and cell number measurements

Mature seeds were used for cell size and cell number measurements. For cell size determination, seeds were imbibed 4 hours in water then overnight in 20% sucrose dissolved in Tris-Buffered Saline solution (TBS, pH 7.5) at 4°C. Sections (20µm) were made by cryodissection at -20°C (Leica). Sections were stained using toluidine blue and cell density was calculated according to a given area to determine the relative size of cells between WT and mutant lines. For cell number determination, imbibed mature seeds were digested for 36h at 37°C in a enzymatic solution containing 0.45M sorbitol, 10 mM MgCl₂, 1mM KH₂PO₄, 20mM MES (pH 5.6), 1% Macerozyme R10 (Phytotechnology), cellulose Onozuka R10 0.4% (phytotechnology). Cell number of a subset of cell suspension was counted using a haemocytometer. This cell number was normalized according to the area of the haemocytometer and to the volume of the digestion solution to predict the total number of cells present in WT and mutant seeds.

II.2.4.7 - Seed clearing

After dissection, seeds from mutants EMS109 and TnT6042 and WT A17 and R108 were fixed in ethanol/acetic acid 9:1 for 1h to overnight. Seeds were then transferred to 90% ethanol for 30 min to 1 h and next incubated for 30min to 1h in 70% ethanol. After these washing steps, seeds were transferred onto microscope slides and covered with a chloral hydrate solution (4g chloral hydrate, 0.5mL glycerol 100%, 1mL water). Samples were sealed with cover slip and transferred to 4°C for clearing during 4 h. Microscopic observations were performed with a Zeiss Axio photomicroscope (Carl Zeiss, Cologne, Germany) equipped with Nomarski differential interference contrast optics.

II.2.4.8 - Light microscopy

Developing seeds (8 -10 dap) from dash and WT genotypes were vacuum-infiltrated overnight at 4°C with a fixative mixture containing 3% (v/v) glutaraldehyde and 2% (w/v) paraformaldehyde in 0.1M sodium phosphate buffered medium (pH7.2). Seeds were then washed and dehydrated in ethanol before embedding in Historesin (Leica, Rueil-Malmaison, France) following provider's instructions. Thick sections (1µm) were cut on a Reichert ultramicrotome (Leica, Rueil-Malmaison, France), then mounted on glass slides and stained with 0.1% (w/v) toluidine blue plus 0.5% (w/v) methylene blue, pH 9, prior to examination by bright field microscopy with a DMRB microscope (Leica, Rueil-Malmaison, France). At least three biological replicates of dash and WT seeds were analyzed at each stage.

II.2.4.9 - Complementation of dash mutant plants

A fragment of 1,150 bp of the native promoter with the 1,011 bp of the coding sequence of *MtDASH* was amplified by PCR using forward primer 5'- TGTAATACTAATGTTTCTTGACTG -3' and the reverse primer 5'- ACAACCTTACTCAACTCAGAACTCC -3'. The pDASH::DASH construct was introduced into the pMDC123 vector (Curtis and Grossniklaus, 2003). The leaves of *M. truncatula* (Jemalong A17) mutant were infected with the *Agrobacterium tumefaciens* EHA105 strain harboring the complementation vector (Cosson et al.,

2006). Resistant calluses obtained after 1-1.5 month of selection were transferred onto regeneration medium. The regenerated shoots or plantlets obtained after 3-4 month were transferred to plastic vessels containing half-strength SH9 medium supplemented with 1mg/L IAA. All the regenerating cultures were kept at 25°C under fluorescent light ($140 \mu\text{E m}^{-2} \text{s}^{-1}$) at a photoperiod of 16 h in the growth chamber. After 3-4 weeks, plants with well-developed roots were transferred to soil and grown in the greenhouse (16 h light, $390 \mu\text{E m}^{-2} \text{s}^{-1}$). Among transformant lines containing the *pDASH::DASH* construct selected based on the presence of the construct and the selection marker, two lines, EMS109 Compl A and EMS109 Compl B displayed complementation of the seed lethal phenotype and were selected for further experiments.

II.2.4.10 - cDNA synthesis and quantitative real-time PCR analysis

About 2 or 3 μg of total RNA were treated with RNase free DNase RQ 1 (Promega, Madison, USA) for 45 min at 37°C according to the manufacturer's instructions. The absence of genomic DNA contamination was confirmed by PCR on non-reverse-transcribed RNA. First-strand cDNA was produced from 2 μg of total RNA using a mix of oligodT and hexamer primers with iScript reverse transcriptase (Bio-Rad, CA, USA) according to the manufacturer's instructions. qRT-PCR was run on LightCycler 480 (Roche) using GoTaq qPCR Master Mix (Promega, Madison, USA). 1 μL cDNA and 200 nM of each gene specific primer in a final volume of 5 μL , were incubated at: 95°C for 5 min; and 50 cycles of 95°C for 10 s and 60°C for 1min. Immediately after thermal cycling, a melting curve was performed to reveal the presence of non-specific amplicons including primer dimerization. The housekeeping genes *ACTIN*, *GAPDH* and *PDF2* were used to normalize candidate gene transcripts (Verdier et al., 2008). The $2^{-\Delta\Delta\text{CT}}$ method was used to calculate relative gene expression (Livak and Schmittgen, 2001). Each expression profile was repeated three times (three biological replicates) in three technical repetitions.

II.2.4.11 - Affymetrix Microarray, Data Extraction, and Normalization

Total RNA was extracted using a modified cetyltrimethylammonium bromide method (Verdier et al., 2008). 5 μg of total RNA from each sample was purified (RNeasy MinElute Cleanup kit; Qiagen) according to the manufacturer's instructions. RNA was quantified and evaluated for purity using an ND-1000 Nanodrop Spectrophotometer (NanoDrop Technologies) and a Bioanalyzer 2100 (Agilent). The Affymetrix *M. truncatula* GeneChip Array (Affymetrix) was used for expression analysis during seed development. RNA from three (for *dash* mutant analysis) independent biological replicates were analysed for each time point. Probe synthesis/labelling was carried out from 500 ng of RNA using the GeneChip 3'IVT express kit, according to the manufacturer's instructions (Affymetrix). Array hybridization, scanning, and data normalization were performed as described by Benedito et al. (2008). Each file from the hybridized Affymetrix array was exported from GeneChip Operating Software version 1.4 (Affymetrix) and imported into Robust Multiarray Average express (Irizarry et al., 2003) for global normalization. Presence/absence call for each probe set to remove background noise was obtained using dCHIP (Li and Wong, 2001).

To identify probesets differentially expressed in dash mutant vs WT control, the R package Anapuce (J. Aubert, UMR 518 AgroParisTech/INRA) was used. For each probeset and dash/wt comparison, a paired t test was performed on the log-2 expression data from three arrays (three independent biological repeats for 8 dap data), assuming that the variance of the log expression was the same for all transcripts per genotype. Spots with extreme specific variance, too low or too high, were excluded from the statistical analysis (details on the procedure given in Gagnot et al. (2008) (Gagnot et al., 2008)). P values were adjusted by the Benjamini-Hochberg method (Benjamini and Hochberg, 1995), which controls the family-wise error rate. Transcripts were considered as differentially represented when associated with a P-value ≤ 0.05 .

The probeset lists of differentially expressed genes were further analysed through Pageman analysis (Usadel et al., 2006) available at <http://mapman.gabipd.org>. MapMan BINs were assigned to each oligonucleotide using the affychip file Mt_AFFY_Mt3.0_0510. Wilcoxon rank sum tests were performed to test whether there were bins that were significantly and consistently behaving different than the other bins in the MapMan ontology using the built-in function in MapMan.

II.2.4.12 - Nimblegen Microarray, Data Extraction, and Normalization

Custom-designed *M. truncatula* Nimblegen arrays were designed based on the *M. truncatula* genome (Young et al., 2011) and containing sequences of the IMGAG Mt3.5.1 version as well as additional genomic sequences (F. Debellé, P. Gamas, and J. Gouzy, personal communication). High-density microarray slides containing 119,780 quality-filtered unigenes were designed and produced by Nimblegen. Total RNA of 12 dap-seed tissues (seed coat, embryo and endosperm) from A17 genotype were extracted using the RNeasy Micro Kit (Qiagen), and 400 ng was amplified using the Ambion messageAmp II following the manufacturer's instructions. Five micrograms of amplified RNA was retrotranscribed with 400 units of SuperScript II reverse transcriptase (Invitrogen), labeled with 1.5 mmol of Cyanine-3 or Cyanine-5 (Interchim), and then purified using NucleoSpin Gel and PCR Clean-up column kits (Macherey-Nagel). Purified and labeled complementary DNA was quantified using NanoDrop ND-1000. 30 pmol of labeled samples were mixed (endosperm with embryo, endosperm with seed coat and embryo with seed coat) and cohybridized to the Medtr_v1.0 12x135K arrays. Three biological replicates were analyzed per comparison using dye switch as described by Churchill et al. (2002) (Churchill, 2002). Hybridizations were performed on a NimbleGen Hybridization System 4 at 42°C overnight. Slides were washed, dried, and scanned at 2- μ m resolution and high sensitivity with a Roche-NimbleGen MS200. NimbleScan Software version 1.2 was used to extract pair data files from the scanned images. Statistical analyses of the gene expression data were performed using the R language, version 2.14, and the Linear Models for Microarray Analysis package (Smyth et al., 2005) from the Bioconductor project. For the preprocessing step, data were normalized by the lowess method. Log ratio and log intensity were calculated before differential expression analyses were

performed using the ImFit function and the Bayes-moderated Benjamini and Hochberg test in Linear Models for Microarray Analysis.

II.2.4.13 - Phytohormone content measurements and auxin treatments

Individual flowers were tagged and developing pods were collected and immediately frozen in liquid nitrogen after harvesting. Pods were grinding using pestle and mortar and 50 mg of tissue powders were used for phytohormones quantification. Phytohormones were extracted using 1mL of extraction solvent (isopropanol:water:HCl at the ratio 2:1:0.002). Internal standards (50 pg) of different hormones to be quantified were added such as d5-IAA, MedhJA, d6ABA, d6SA and 13cBA. Extractions were performed by shaking samples at 4°C for 1h, then 0.5mL of dichloromethane was added and samples shaken for another 30 min. After centrifugation at 3000rpm (30 min at 4°C), bottom layers were collected and dried under nitrogen. Dry pellets were resuspended in 0.1mL methanol and 1mL of acetic acid 1% then cleaned using conditioned Waters Oasis HLB columns according to manufacturer's instructions (Waters). Hormone samples were eluted from the columns using 1.8mL of 80% methanol containing 1% acetic acid to glass vials, dried under nitrogen. Dry samples were dissolved in 25µL methanol and 25 µL 1% acetic acid and 10 µL was injected in an Agilent triple quadrupole LC/MSMS analyser. A reverse phase 1.7µm UPLC BEH C18 (2.1x150mm) column was used for separation. Data were analysed using MassHunter quantitative analysis software (Agilent).

Individual flowers (50 to 100) were tagged and developing pods were treated twice at 6 and 8 dap for 3 plants per condition. Treatments consisted on pods immersion few seconds in water, AIA or TIBA solution (concentration of 100 mg.L⁻¹ supplemented with 1 drop of Tween for 50mL solution). Each treatment (water or hormone) was repeated on three WT or EMS plants. Pods were harvested at maturity weight measurement was performed on about 20 pods containing-seeds and isolated seeds per plant and per treatment.

Acknowledgements

We would like to thank Peter Rogowsky and Bertrand Dubreucq for helpful advices. Lloyd Sumner and David Huhman for their help for phytohormone content measurement. Zhenbiao Yang for his critical review of this manuscript.

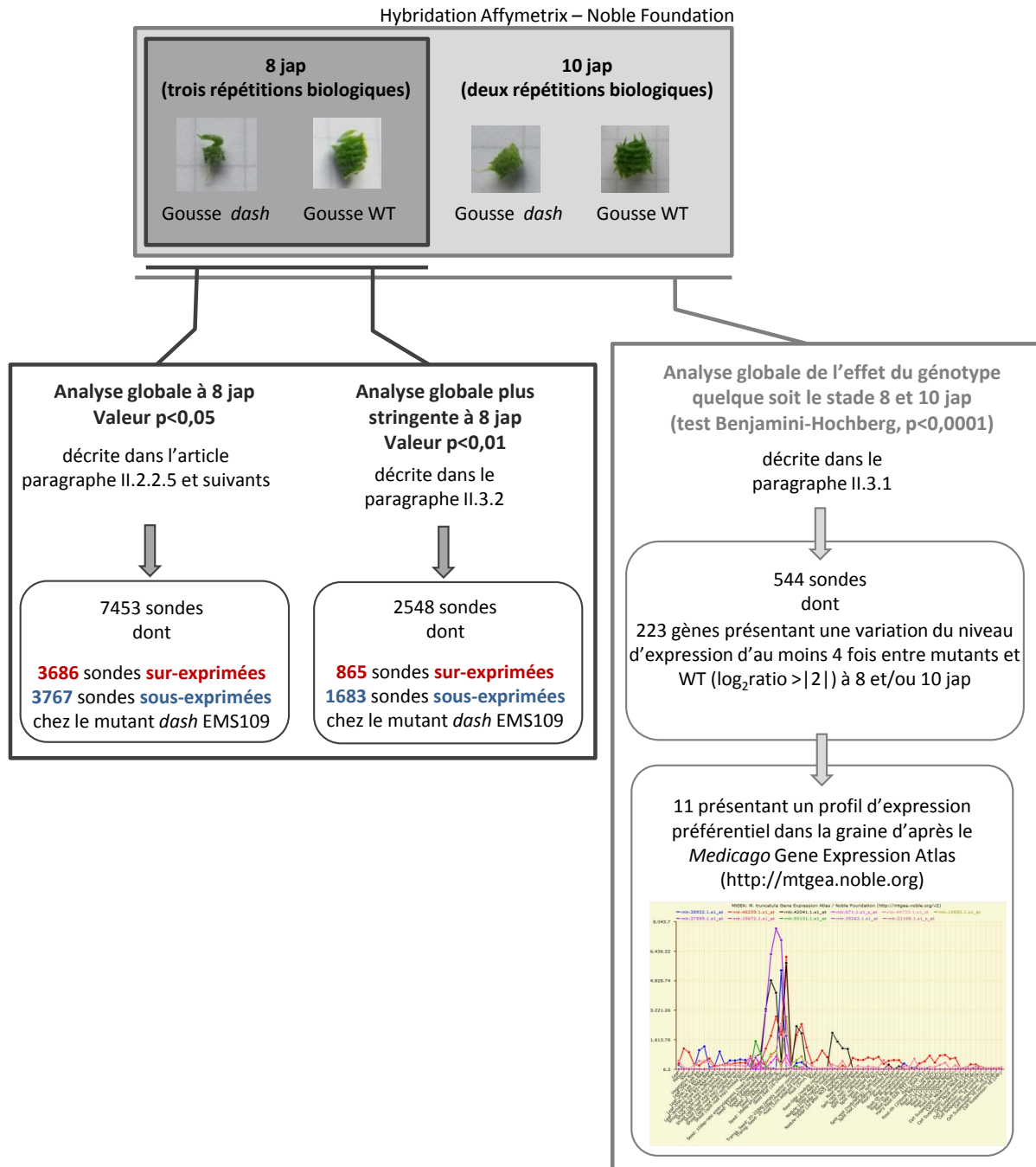


Figure II-8 – Bilan des différentes approches utilisées pour l'analyse comparative du transcriptome de gousses du mutant *dash* EMS109 vs WT à 8 et 10 jap.

II.3 - Données complémentaires à l'article : recherche de cibles potentielles de DASH via une approche transcriptomique

Plusieurs approches ont été exploitées afin d'analyser les données transcriptomiques précédemment décrites dans l'article. Ces différentes stratégies d'analyses, présentées dans la figure II-8, ont été développées avant d'opter pour l'approche finale décrite dans l'article et sont présentées dans cette partie.

La première approche (décrite dans le paragraphe suivant II.3.1) a été conduite afin de sélectionner des cibles directes du gène *DASH*, c'est-à-dire les gènes dérégulés chez le mutant *dash* EMS 109 à deux stades de développement 8 jap (trois répétitions biologiques) et à 10 jap (deux répétitions biologiques). Puis, pour des raisons de fiabilité et de publication des résultats, les données à 10 jap ne comportant que deux répétitions biologiques ont été écartées. Deux valeurs seuils de probabilité ont été appliquées afin de s'intéresser aux gènes dérégulés chez le mutant *dash* EMS 109 à 8 jap : analyse avec valeur $p < 0,05$ décrite dans l'article et analyse avec valeur $p < 0,01$ présentée dans la partie II.3.2. Cette valeur seuil $p < 0,01$ étant plus stringente, cette dernière approche nous a permis proposer des gènes cibles possibles de DASH en s'affranchissant des voies métaboliques dérégulées de façon indirecte.

II.3.1 – Analyse globale de l'effet génotype quelque soit le stade 8 et 10 jap

Cette première stratégie abordée est présentée dans la figure II-8. Elle a été envisagée en vue d'identifier les gènes susceptibles d'être des cibles du facteur de transcription DASH. Pour cela, les données d'expression générées dans les gousses à 8 jap (trois répétitions biologiques) et à 10 jap (deux répétitions biologiques) ont été prises en compte. Une analyse de variance à deux facteurs (génotype et stade) suivi d'un test de Benjamini-Hochberg (analyses statistiques réalisées par Christine Le Signor) a permis de sélectionner 544 sondes ayant des profils d'expression qui varient significativement en fonction du génotype aux deux stades étudiés. Nous avons ensuite sélectionné les sondes dont le signal variait entre le sauvage et le mutant d'au moins quatre fois ($\log_2 \text{ratio} > |2|$), soit 223 sondes [figure II-8]. Il est à noter que l'on retrouve dans cette sélection une part importante de gènes (199) au moins quatre fois plus exprimés chez le mutant à 8 et/ou 10 jap, suggérant ainsi que *DASH* serait plutôt un represseur. Les profils d'expression de ces 223 sondes ont été analysés via le *Medicago truncatula* Gene Expression Atlas (MtGEA) (Benedito et al., 2008; He et al., 2009) afin de déterminer si les sondes présentent un profil d'expression semblable à celui du gène *DASH* (cf figure I-1, article). Ceci a permis d'identifier 11 gènes

Medicago truncatula						Glycine max							Arabidopsis thaliana						
identifiant Affymetrix	identifiant Mt3.5	description	code Bin et description	log ₂ ratio EMS/A17 at 8jap	log ₂ ratio EMS/A17 at 10jap	Orthologue putatif (tblastx)		GS	HS	CS	EMS	Orthologue putatif (tblastx)		GS	HS	CS	MS		
mtr.38932.1.s1_at	AC233577_31.1	Malate synthase (cycle glyoxylate)	6.2 - gluconeogenese/ glyoxylate cycle.malate synthase	6,58	2,56	GmaAffx.91133.1.S1_s_at	malate synthase	teg alb emb	teg emb	teg emb	teg emb	At5g03860	malate synthase	-	-	-	alb emb		
mtr.46259.1.s1_at	Medtr2g063610.1	Metacaspase type II	29.5 - protein.degradation	4,83	3,02	Gma.7051.1.A1_at	metacaspase-9-like	-	alb	teg	-	At5g04200	metacaspase 9 AtMC9	teg	teg	teg alb	teg alb		
mtr.42041.1.s1_at	ND		33.99 - development.unspecified	3,93	1,76	ND						ND							
mtr.671.1.s1_s_at	Medtr1g102390.1	Cytochrome P450 704C1	26.10 - misc.cytochrome P450	3,78	1,40	GmaAffx.28428.1.S1_at	cytochrome P450 CYP704C1-like	pas d'expression				AT2G45510	cytochrome P450, CYP705A2	-	teg alb emb	teg alb emb	alb emb		
mtr.44733.1.s1_at	Medtr3g046650.1	hypothetical protein	35.2 - not assigned.unknown	3,01	0,77	ND						ND							
mtr.10693.1.s1_at	Medtr7g075540.1*	Medicago truncatula Cytochrome P450	26.10 - misc.cytochrome P450	2,33	0,81	Gma.2227.1.S1_at	cytochrome P450 CYP706A4	emb	alb emb	emb	emb	At2g45550	cytochrome P450 CYP76C7	pas d'expression					
mtr.37595.1.s1_at	Medtr4g060780.1*	Medicago truncatula Allergen Gly m Bd	33.1 - development.storage proteins	2,10	0,21	Gma.13356.1.S1_at	allergen Gly m Bd 28K	-	-	-	emb	At2g28490	putative seed storage protein (vicilin)	-	-	-	teg alb emb		
mtr.15672.1.s1_at	Medtr1g056330.1 *	Disease resistance protein	35.2 - not assigned.unknown	-0,13	-3,10	ND						At4g27200	pseudogene/transposon	pas d'expression					
mtr.50101.1.s1_at	Medtr6g009690.1**	Unknown Protein	35.2 - not assigned.unknown	-5,75	-6,37	ND						ND							
mtr.35262.1.s1_at	Medtr6g055570.1**	Subtilisin-like serine protease	29.5.1 - protein.degradation.subtilases	-1,25	-4,23	GmaAffx.68175.1.S1_at	subtilisin-like protease C1	-	-	-	emb	At5g59190	subtilase family protein	alb	-	-	-		
mtr.22188.1.s1_s_at	Medtr5g088180.1**	Lamin-like protein	26.19 - misc.plastocyanin-like	-1,01	-3,26	ND						At5g15350	putative protein many predicted proteins	teg alb emb	teg alb emb	teg alb emb	alb emb		

*site d'expression dans les tissus de la graine d'après les données disponibles sur <http://seedgenenetwork.net/soybean>

*site d'expression dans les tissus de la graine d'après les données disponibles sur <http://seedgenenetwork.net/arabidopsis>

abréviations: alb: albumen, emb: embryon, teg:tégument, GS: stade globulaire, HS: stade cœur, CS: stade cotylédons, EMS: stade début maturation, MS: stade mature.

moyenne des intensités des signaux (<http://seedgenenetwork.net/soybean>)



Tableau II-3 – Liste des 11 gènes les plus fortement sur ou sous exprimés chez le mutant *dash* EMS109 par rapport au WT quelque soit le stade (8 et 10 jap), et préférentiellement exprimés dans la graine. Les orthologues chez le soja et *A. thaliana* ont été recherchés ainsi que leur site et niveau d'expression dans les tissus de la graine.

* gènes significativement dérégulés suite à l'analyse transcriptomiques à 8 jap ($p < 0,05$) (cf données transcriptomiques article)

** gènes significativement dérégulés suite à l'analyse transcriptomique à 8 jap ($p < 0,01$) et ($p < 0,05$)

préférentiellement exprimés dans la graine au stade 12 jap, dont la liste est présentée dans le tableau II-3. Les orthologues de ces gènes chez *A. thaliana* et le soja, ainsi que leur site et niveau d'expression dans les tissus de la graine (données Harada et Goldberg disponibles sur le portail <http://seedgenenetwork.net/>) ont été recherchés. Il est à noter que l'on retrouve le gène Medtr6g009690.1 (identifiant version Mt3.5), précédemment identifié dans l'analyse transcriptomique à 8 jap présentée dans l'article en préparation. Ce gène est présent parmi les plus dérégulés (cf figure II-6) mais aucun gène orthologue n'a cependant pu être identifié pour ce gène. Un autre gène parmi la liste code une protéase à sérine de type subtilisine (Medtr6g055570.1, identifiant version Mt3.5). Une étude précédente conduite au laboratoire a révélé qu'une subtilisine associée à l'albumen des graines (Medtr5g016120.1, identifiant version Mt3.5) serait impliquée dans le contrôle de la taille de la graine chez *M. truncatula* et le pois (D'Erfurth et al., 2012). Bien que cette protéine ne présente que 38% d'identité de séquence avec la protéase à sérine identifiée dans la présente étude, il sera intéressant de tester si ces protéases à sérines sont des cibles de DASH. Il est à souligner qu'il n'a pas toujours été possible d'identifier des gènes orthologues chez *A. thaliana* ou le soja, et parmi les orthologues putatifs proposés, tous ne s'expriment pas dans la graine. Aucun mutant pour ces gènes chez *A. thaliana* ou le soja n'a été décrit comme induisant un phénotype particulier dans le développement de la graine. Il est donc possible que ces gènes aient des fonctions différentes entre espèces. L'objectif étant ici d'identifier des cibles potentielles de DASH, la présence du motif "AAAG" décrit comme nécessaire à la fixation des TFs de type DOF (Yanagisawa, 2002; Umemura et al., 2004; Benedito et al., 2008; He et al., 2009) a été vérifiée dans les séquences promotrices de ces 11 gènes. La recherche sur 2 500 à 3 000 pb en amont de l'ATG a permis d'identifier de nombreuses fois le motif "AAAG" mais étant donné la longueur très courte de ce motif, il est impossible de déterminer si leur présence est aléatoire ou liée à un site de fixation à un TF de type DOF.

II.3.2 – Analyse globale plus stringente ($p < 0,01$) des données transcriptomiques à 8 jap

Une seconde analyse des données via SAS (Statistical Analysis System), réalisée par Christine Le Signor et présentée dans l'article, a permis d'identifier les sondes dont l'expression variait significativement chez le mutant, comparé au génotype sauvage, à 8 jap uniquement. A ce stade, les mutants *dash* EMS109 ne présentent pas encore de modification drastique au niveau de la morphologie de l'embryon bien que l'expression de *DASH* augmente significativement (cf. figures II-1 et II-5). Cette analyse permet donc de s'affranchir des changements massifs liés au retard de croissance de l'embryon en vue de sélectionner des cibles possibles directes de DASH. Les données relatives à la sélection de gènes avec une valeur p inférieure à 0,05 ont été présentées dans l'article et rendent compte des changements globaux intervenant dans le transcriptome du mutant *dash* (EMS109). Les données présentées dans cette partie ont été sélectionnées avec une valeur p plus stringente dans une perspective d'identification de gènes cibles. Cette analyse nous a permis de sélectionner 2548 sondes montrant une différence significative avec un seuil de valeur $p < 0,01$.

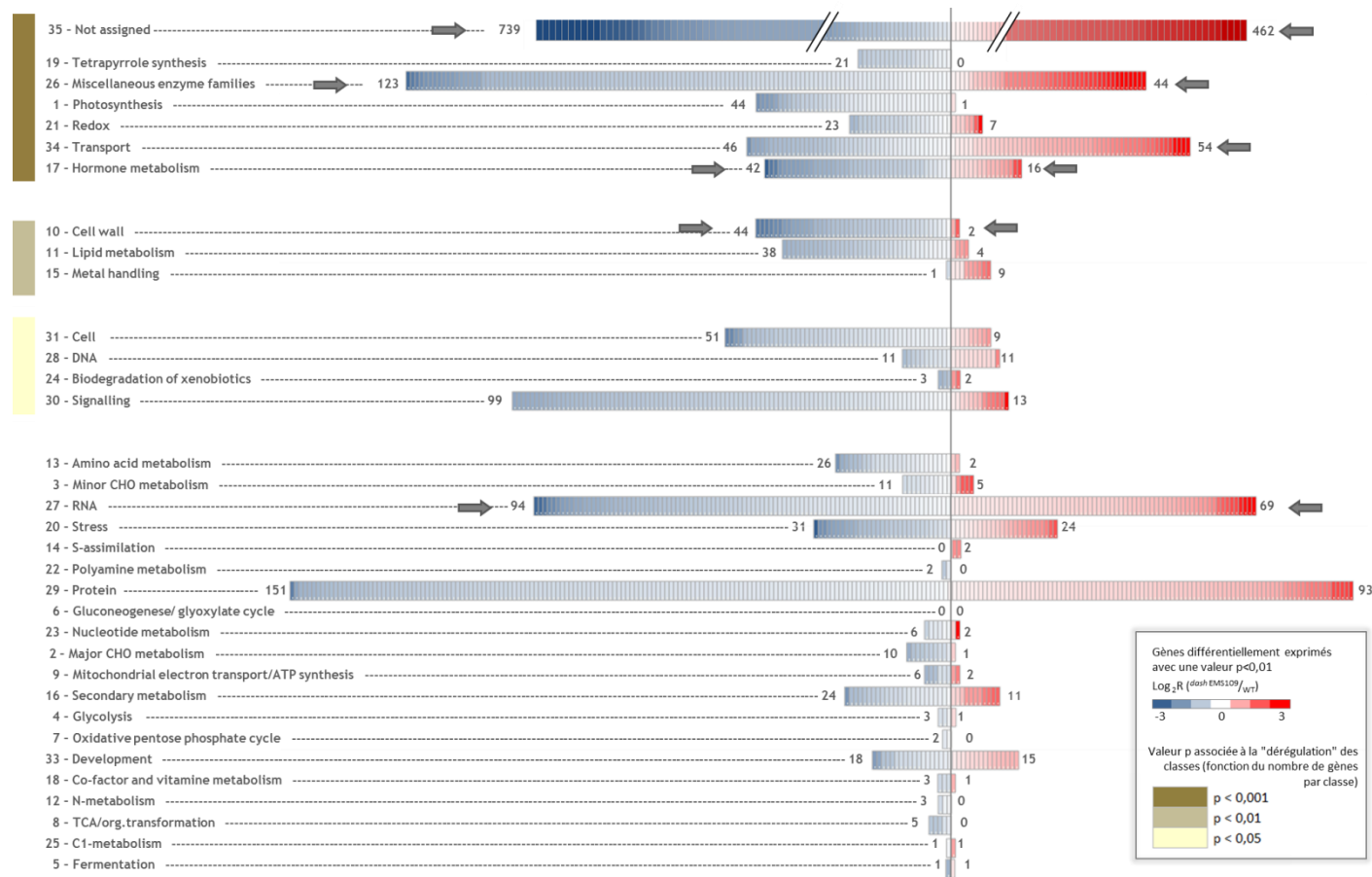


Figure II-9 – Répartition des gènes dérégulés ($p < 0,01$) en fonction des classes fonctionnelles BIN avec en bleu les gènes sous-exprimés chez le mutant *dash* EMS109 et en rouge les gènes sur-exprimés, l'intensité des couleurs est relative à l'intensité de sur- ou sous-expression. Les valeurs indiquent le nombre de gènes pour chaque classe fonctionnelle. Sur la gauche, le code couleur correspond aux valeurs p (test de Fisher exact) relatives à la significativité de la dérégulation du nombre de gènes par classe fonctionnelle par rapport à la totalité des gènes de chaque classe présents sur la puce. Les flèches indiquent que plusieurs gènes de ces classes parmi les plus fortement dérégulés ont été sélectionnés en vue de confirmer les profils de sur- ou sous-expression par q-PCR.

Parmi les 2548 gènes sélectionnés, 865 montrent une sur-expression chez le mutant *dash* (jusqu'à 148 fois supérieure pour le mutant EMS109 par rapport au type sauvage) et 1683 sont sous-exprimés jusqu'à 98 fois chez le mutant. Trois gènes précédemment identifiés *via* l'analyse des données à 8 et 10 jap (tableau II-3) figurent dans cette liste, dont le gène Medtr6g055570.1 codant la protéase à sérine de type subtilisine et le gène Medtr6g009690.1 ne présentant pas d'orthologues défini. Afin d'interpréter en parallèle les résultats sous forme de classes fonctionnelles, les identifiants dans la version Mt3.5 et les codes BIN (*i.e.* classes fonctionnelles) selon MapMan (Thimm et al., 2004) ont été recherchés.

Nous avons ensuite comparé la répartition de ces gènes dans les classes fonctionnelles avec la répartition de ces classes pour l'ensemble des gènes de la puce. Une première analyse a été réalisée en utilisant le test de Fisher exact. La figure II-9 reprend l'ensemble de ces données, avec les classes fonctionnelles ordonnées en fonction des valeurs p associées à leur "dérégulation" (dégradé de jaune au marron à gauche de la figure), c'est-à-dire la comparaison entre la répartition du nombre de gènes par classe fonctionnelle pour l'ensemble des gènes qui ressortent différentiellement exprimés avec une valeur $p < 0,01$ par rapport au nombre de gènes de chaque classe présents sur la puce [annexe II-7]. Les classes les plus dérégulées apparaissent ainsi sur cette figure qui indique également le niveau de sur- ou sous-expression des gènes de chaque classe, avec en rouge les gènes sur-exprimés chez le mutant et en bleu les gènes sous-exprimés. On remarque que certaines classes : "35.not assigned", "17.hormone metabolism", "26.miscellaneous enzyme families" et "27.RNA" contiennent des sondes à la fois fortement sur- et sous-exprimées. A l'inverse, d'autres classes : "1.photosynthesis", "19.tetrapyrrole synthesis", "11.lipid metabolism" ou "15.metal handling" sont fortement dérégulées mais les gènes présents dans ces classes ne sont pas fortement sur- ou sous-exprimés ($\log_2 R < |1,5|$).


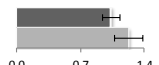
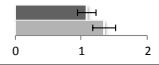
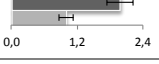
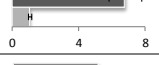
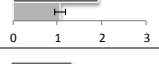
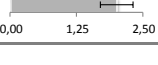
Une analyse de ces données *via* MapMan (Thimm et al., 2004) [annexe II-8] a été réalisée en parallèle, mettant en évidence les classes fonctionnelles et le nombre de gènes qu'elles comportent avec des dégradés de couleur représentatifs de l'intensité de la dérégulation (dégradé de rouge pour les gènes les plus sur-exprimés et dégradé de bleu pour les plus sous-exprimés chez le mutant *dash* EMS109). Trois classes fonctionnelles se distinguent par un nombre de gènes dérégulés plus important que les autres. Il s'agit des classes "10.cell wall" qui comporte un nombre important de gènes fortement sous-exprimés, "17.hormone metabolism" qui est également affectée par de nombreux gènes sous-exprimés et seulement deux fortement sur-exprimés et "26.miscellaneous enzyme family" où les gènes fortement sur-exprimés sont présents en nombre. Ces mêmes classes ont également été identifiées dans la première analyse *via* le test de Fisher exact [figure II-9].

II.3.2.1 - Validation du différentiel d'expression de gènes cibles potentiels de DASH

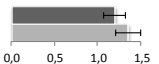
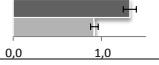

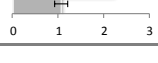
Parmi les classes fonctionnelles les plus affectées (indiquées par des flèches sur la figure II-9), plusieurs gènes ont été sélectionnés parmi les plus fortement sur- ou sous- exprimés chez le mutant *dash* (cinq à dix pour chaque classe, $\log_2 R > |1,5|$). Il est à noter qu'il n'a pas toujours été possible d'identifier les gènes

Identifiant Affimetrix	valeur p	log ₂ ratio dash/WT	code BIN	description BIN	identifiant Mt3.5.1	description Mt3.5	Valeurs d'expression relative des gènes (qPCR)			Expression dans la graine d'après le GeneAtlas	Expression dans les tissus de la graine à 12 jap (valeur en log ₂)			Expression préférentielle graine/gousse (qPCR)
							EMS	WT			Alb	Teg	Emb	

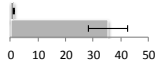
Classes fonctionnelles 17.hormone metabolism et 27.3. RNA.regulation of transcription

Mtr.43023.1.S1_at	0,0023	2,18	17.2.3	hormone metabolism.auxin.induced-regulated-responsive-activated	Medtr7g021850.1	aldo/keto reductase family protein	3,88 ± 1,18	1,08 ± 0,13		oui mais non spécifique	5,23	7,67	3,83	graine
Mtr.20620.1.S1_x_at	0,0034	-1,80	17.2.3	hormone metabolism.auxin.induced-regulated-responsive-activated	Medtr3g084230.1	Auxin-induced protein-like	1,23 ± 0,15	1,04 ± 0,09		pas d'expression	0,58	1,06	0,62	gousse
Mtr.49768.1.S1_x_at	0,0047	-2,39	17.2.3	hormone metabolism.auxin.induced-regulated-responsive-activated	Medtr4g072240.1	Auxin-induced protein 6B	1,35 ± 0,17	1,09 ± 0,14		pas d'expression	0,28	2,16	0,30	gousse
Mtr.49793.1.S1_x_at	0,0093	-2,50	17.2.3	hormone metabolism.auxin.induced-regulated-responsive-activated	Medtr4g072610.1	Auxin-induced SAUR-like protein	1,00 ± 0,13	1,83 ± 0,24		pas d'expression	1,05	1,35	1,13	gousse
Mtr.46809.1.S1_at	0,0002	-1,82	17.7.1.2	hormone metabolism.jasmonate.synthesis-degradation.lipoxygenase	Medtr8g018430.1	Lipoxygenase	1,06 ± 0,13	6,80 ± 0,98		oui mais non spécifique	0,22	1,95	0,05	graine
Mtr.43054.1.S1_at	2E-05	-2,19	27.3.40	RNA.regulation of transcription.Aux/IAA family	Medtr1g093240.1	Indoleacetic acid-induced-like protein	1,06 ± 0,12	1,91 ± 0,15		pas d'expression	3,89	3,35	3,56	gousse
Mtr.45730.1.S1_at	0,0095	3,99	27.3.3	RNA.regulation of transcription.AP2/EREBP	Medtr1g019110.1	Ethylene-responsive transcription factor 5	2,01 ± 0,30	1,16 ± 0,20		oui mais non spécifique	2,15	5,93	1,25	ND

Classes fonctionnelles 10.cell wall

Mtr.48647.1.S1_s_at	0,0003	-2,25	10.2.1	cell wall.cellulose synthesis.cellulose synthase	Medtr2g087900.1	Cellulose synthase	1,35 ± 0,14	1,20 ± 0,13		pas d'expression	-0,39	-0,24	3,05	ND
Mtr.47442.1.S1_s_at	0,0071	-1,87	10.6.2	cell wall.degradation.mannan-xylose-arabinose-fucose	Medtr7g109390.1	Beta xylosidase	0,92 ± 0,04	1,31 ± 0,08		oui mais non spécifique	1,14	6,05	3,09	gousse
Mtr.26579.1.S1_s_at	0,0001	-2,44	10.6.3	cell wall.degradation.pectate lyases and polygalacturonases	Medtr7g088550.1	Poly-galacturonase	1,21 ± 0,25	14,90 ± 2,76		spécifique graine	4,93	1,88	1,27	graine
Mtr.16137.1.S1_at	0,0002	-2,38	10.7	cell wall.modification	Medtr2g095800.1	Xyloglucan endotraglucosylase/hydrolase	1,06 ± 0,14	2,25 ± 0,21		oui mais non spécifique	0,43	1,31	0,21	graine

Classe fonctionnelle 34.transport

Mtr.25128.1.S1_at	0,0074	4,21	34.99	transport.misc	Medtr5g032720.1	antiporter/ drug transporter	35,31 ± 7,00	1,167 ± 0,214		pas d'expression	3,95	3,04	4,38	graine et gousse
-------------------	--------	------	-------	----------------	-----------------	------------------------------	--------------	---------------	---	------------------	------	------	------	------------------

CHAPITRE -2-

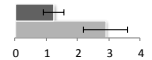
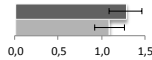
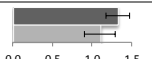
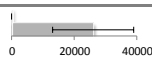
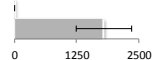

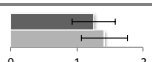
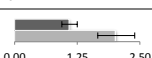
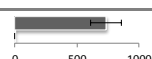

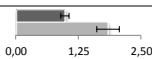
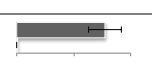
Identifiant Affymetrix	valeur p	log ₂ ratio dash/WT	code BIN	description BIN	identifiant Mt3.5.1	description Mt3.5	Valeurs d'expression relative des gènes (qPCR)		Expression dans la graine d'après le GeneAtlas	Expression dans les tissus de la graine à 12 jap (valeur en log ₂)			Expression préférentielle graine/gousse (qPCR)	
							EMS	WT		Alb	Teg	Emb		
Classe fonctionnelle 26.miscellaneous enzyme family														
Mtr.13532.1.S1_at	0,0041	2,81	26.10	misc.cytochrome P450	Medtr5g016440.1	CYP81D1 electron carrier/ heme binding / iron ion binding / monooxygenase/ oxygen binding	2,89 ± 0,72	1,24 ± 0,34		pas d'expression	-0,27	-0,32	-0,29	graine et gousse
Mtr.43628.1.S1_at	0,0015	-1,95	26.2	misc.UDP glucosyl and glucoronyl transferases	Medtr5g067170.1	UDP-glucosyltransferase	1,09 ± 0,17	1,28 ± 0,19		oui mais non spécifique	0,64	0,65	0,65	gousse
Mtr.43075.1.S1_at	0	-2,73	26.3	misc.gluco-, galacto- and mannosidases	Medtr3g026620.1	Beta-glucosidase 12	1,10 ± 0,20	1,33 ± 0,15		oui mais non spécifique	0,48	1,79	5,74	gousse
Classe fonctionnelle 26.miscellaneous enzyme family														
Mtr.37279.1.S1_at	0,0008	7,22	26.22	misc.short chain dehydrogenase/reductase (SDR)	Medtr6g023610.1	Short-chain alcohol dehydrogenase	26025 ± 12881	1,21 ± 0,26		pas d'expression	0,68	0,59	0,69	graine
Mtr.33537.1.S1_at	0,0067	2,94	26.22	misc.short chain dehydrogenase/reductase (SDR)	Medtr6g023940.1	short-chain dehydrogenase/reductase SDR family protein	1792 ± 557	1,03 ± 0,08		pas d'expression	1,04	1,34	0,34	graine et gousse
Mtr.40558.1.S1_at	0,0087	4,11	26.22	misc.short chain dehydrogenase/reductase (SDR)	Medtr6g023910.1	short-chain dehydrogenase/reductase SDR family protein	306,6 ± 126,0	6,02 ± 3,83		pas d'expression	0,80	1,34	0,12	graine
Classe fonctionnelle 26.miscellaneous enzyme family														
Msa.1843.1.S1_at	0,0041	-1,96	26.28	misc.GDSL-motif lipase	Medtr7g116520.1	GDSL esterase/lipase	1,42 ± 0,34	1,25 ± 0,32		pas d'expression	1,18	0,04	0,52	ND
Mtr.46059.1.S1_at	0,0001	-1,99	26.28	misc.GDSL-motif lipase	Medtr1g025910.1	GDSL esterase/lipase	2,02 ± 0,37	1,09 ± 0,15		spécifique graine	1,30	0,22	-0,06	ND
Classe fonctionnelle 35.not assigned														
Mtr.1254.1.S1_at	0,0005	-5,39	35.2	not assigned.unknown	Medtr2g076980.1	Unknown Protein	0,452 ± 0,08	736 ± 126,7		spécifique graine	5,85	5,12	3,59	graine
Mtr.21576.1.S1_at	0,0008	-6,22	35.2	not assigned.unknown	Medtr8g046040.1	Unknown Protein	1,060 ± 0,481	2747 ± 597		spécifique graine	6,86	4,71	3,05	graine
Mtr.4439.1.S1_s_at	0,0013	3,81	35.2	not assigned.unknown	Medtr5g037890.1	Thioredoxin-2	1,84 ± 0,23	0,98 ± 0,09		pas d'expression	1,32	4,98	0,62	ND
Mtr.50101.1.S1_at	2E-05	-5,75	35.2	not assigned.unknown	Medtr6g009690.1	Unknown Protein	1,88 ± 0,54	19434 ± 3608		spécifique graine	5,27	4,72	2,97	ND

Tableau II-4 – Liste des sondes sélectionnées en vue de vérifier leurs variations de niveau d’expression dans la graine. Pour chaque sonde sont indiqués la valeur p de probabilité que les variations d’expression soient dues au génotype, le log₂ du ratio mutant/WT, le code et nom BIN, l’identifiant du gène dans la version Mt3.5 et la description, les données d’expression obtenues par qPCR sur des échantillons de graines à 8 jap, l’expression dans la graine d’après les données du Gene Atlas ainsi que les données d’expression dans les tissus de la graine à 12 jap (données NimbleGen – chapitre 3). La dernière colonne indique l'expression préférentielle dans les graines ou gousses déterminée par q-PCR (génotype A17, peu de répétitions).

correspondant aux sondes, ce qui a réduit la liste des gènes sélectionnés à 24. Les variations de différentiel d'expression pour ces gènes ont ensuite été vérifiées par q-PCR sur des échantillons de graines isolées à 8 jap. Le tableau II-4 présente la liste de ces sondes, la probabilité que les variations d'expression soient dues au génotype, le \log_2 du ratio mutant *dash* EMS109/WT, le code BIN et son nom. L'identifiant du gène dans la version Mt3.5 et sa description, les données d'expression obtenues par qPCR sur des échantillons de graines isolées à 8 jap, l'expression dans la graine d'après les données du Gene Atlas ainsi que les données d'expression dans les tissus de la graine à 12 jap (données NimbleGen – chapitre 3) figurent également dans le tableau. L'expression préférentielle dans la graine ou la gousse est indiquée mais ces résultats sont à interpréter avec réserve (peu de répétitions).

Sur les 24 gènes testés par q-PCR, 17 montrent des profils d'expression similaires à ceux détectés sur la puce. Les hybridations ayant été réalisées avec des tissus de gousses entières, il est important de constater que les variations d'expression sont, dans la majorité des cas validées par q-PCR sur des échantillons de graines isolées.

a - Gènes relatifs au métabolisme hormonal

La classe "17.hormone metabolism" ressort comme une classe fortement dérégulée. Il semblait donc judicieux de valider les profils d'expression de quelques-uns de ces gènes [figure II-9]. Deux gènes de la classe "27.RNA" ont également été sélectionnés car liés au métabolisme hormonal (un FT de la famille AP2/EREB (ethylene response factor) et un FT de la famille Aux/IAA (IAA induced-like protein)). Il est cependant difficile d'interpréter les variations car comme précédemment observé sur la figure II-9, certains gènes de réponse à l'auxine sont sur-exprimés chez le mutant *dash* EMS109 alors que d'autres sont sous-exprimés. Les résultats q-PCR sur des échantillons de graines isolées permettent de confirmer les gènes Medtr4g072610.1, Medtr1g093240.1 sous-exprimés et le gène Medtr7g021850.1 sur-exprimé, certains ne présentant pas de variations (Medtr3g084230.1, Medtr4g072240.1). On peut supposer que les différences entre les résultats d'hybridation de la puce et de q-PCR pourraient être dues à des variations observées au niveau de la gousse plutôt que dans la graine étant donné que la plupart de ces gènes sont préférentiellement exprimés dans la gousse. Le métabolisme hormonal semblait majoritairement sous-régulé [figure II-9] et ce sont particulièrement les gènes de réponse à l'auxine qui sont sous-exprimés [tableau supplémentaire 1, article] et non les gènes relatifs à la synthèse ou dégradation d'auxine, ou encore les gènes de transduction du signal ; ce qui laisse supposer un dysfonctionnement dans le transport de l'auxine vers l'embryon ou la perception de l'auxine au niveau des graines et/ou des gousses. On note également que ces gènes de réponse sont préférentiellement exprimés dans le tégument.

b - Gènes relatifs au métabolisme des sucres

Les gènes impliqués dans la structure des parois cellulaires des classes "10.cell wall" et "26.miscellaneous enzyme families" sont fortement dérégulés [figure II-9]. La validation des différences de

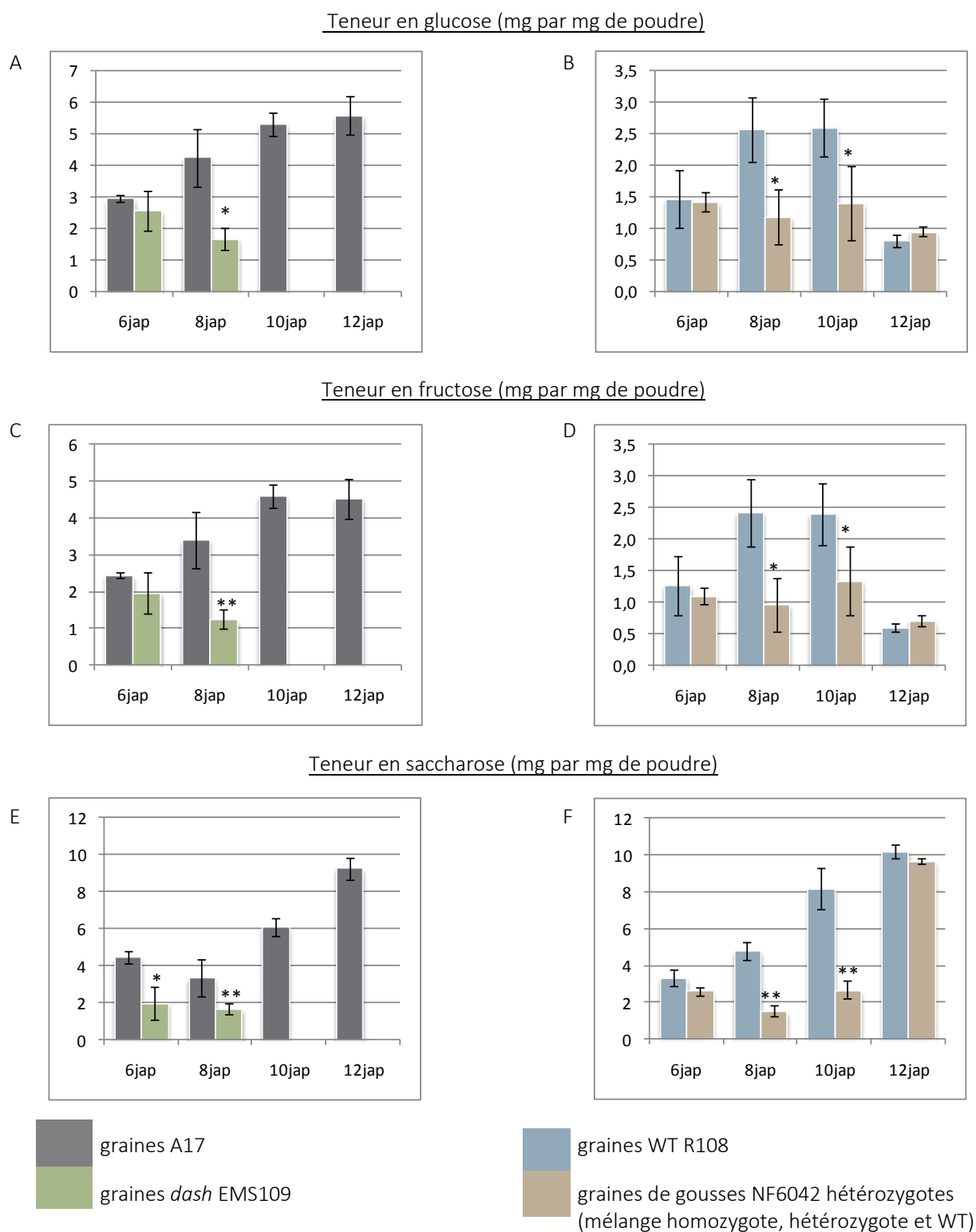


Figure II-10 – Dosage de glucose (A, B), fructose (C, D) et saccharose (E, F) exprimé en mg par mg de poudre de graine lyophilisée, dans les graines en cours de développement chez les mutants *dash* EMS109 (A, C, E) et NF6042 (B, D, F). Ces teneurs en sucre ont été estimées sur 4 à 5 lots de graines différents. * $p < 0,05$, ** $p < 0,01$.

profils d'expression par q-PCR a permis de confirmer trois gènes sur six, la moitié ne montrant pas de variations significatives. Il est intéressant de noter que l'un de ces gènes, appartenant à la famille des polygalacturonases (Medtr7g088550.1), est fortement sous-exprimé chez le mutant *dash* EMS109. Ce gène est très fortement exprimé dans l'albumen [tableau II-4] et présente un profil d'expression spécifique de la graine, d'après les données du Gene Atlas MtGEA (Benedito et al., 2008; He et al., 2009) (également confirmé par qPCR). Ce gène pourrait donc être une cible directe de *DASH*. Outre les besoins en polysaccharides pour la composition des parois cellulaires, la graine a besoin d'un apport en sucres C6 (glucose et fructose) et C12 (sucrose) pour le développement de l'embryon notamment et ces flux d'apport glycosidiques en C12 et C6 transitent via l'albumen (Weber et al., 1997; Hill et al., 2003). Un lien entre la dérégulation du métabolisme hormonal et la composition en sucres de la graine est également envisageable, puisqu'il a déjà été démontré que la balance des sucres C6/C12 pouvait influencer l'expression des gènes YUCCA qui interviennent dans la biosynthèse d'auxine (LeClere et al., 2008; Mishra et al., 2009; LeClere et al., 2010). Des dosages de sucres en C6 (glucose et fructose) et C12 (saccharose) ont été réalisés dans des graines en cours de développement chez les mutants *dash* (EMS109 et NF6042) et leurs génotypes sauvages respectifs (génotypes A17 et R108) [figure II-10]. Ils ont révélé des concentrations en sucres plus faibles dans les graines du mutant *dash* (EMS109) à 8 jap, soit avant qu'elles avortent [figure II-10A, B et C]. Il semblerait donc qu'il y ait un défaut d'accumulation des sucres dans la graine, potentiellement dû à des troubles aux niveaux du transport et/ou de la conversion des sucres ou à un défaut d'utilisation des sucres par les tissus de la graine.

Quant au mutant *dash* NF6042 hétérozygote, les mesures suggèrent également un défaut d'accumulation à 8 et 10 jap [figure II-10D, E et F]. On observe cependant une restauration de la concentration des sucres à 12 jap, ceci pouvant s'expliquer par l'avortement des graines homozygotes contenues dans les gousses hétérozygotes. Les graines soumises à l'analyse des sucres à ce stade sont donc de type hétérozygotes et WT avec une concentration en sucres comparable aux graines d'une plante WT.

Les graines des mutants *dash* EMS109 et NF6042 seraient donc déficientes en sucres à partir de 8 jap, stade associé à l'augmentation de l'expression du gène *DASH*. Il semble donc y avoir un lien entre les teneurs en sucres de la graine et le gène *DASH*, et on suppose que les embryons qui ne reçoivent pas une teneur en sucres suffisante sont bloqués dans leur développement. Afin de tester cette hypothèse il conviendrait de mesurer les concentrations en sucres des graines du mutant *dash* EMS109 produites en fin de cycle pour vérifier que les teneurs en sucres sont rétablies dans ces graines.

c - Gènes de la classe "26.miscellaneous enzyme families"

Parmi les gènes de cette classe, trois sont fortement sur-exprimés chez le mutant d'après les données transcriptomiques et q-PCR entre les graines des génotypes mutant *dash* EMS109 et sauvage. Ces trois gènes (Medtr6g023610.1, Medtr6g023940.1 et Medtr6g023910.1) codent des SDR (short chain dehydrogenase/reductase) et semblent peu exprimés dans la graine d'après les données du Gene Atlas

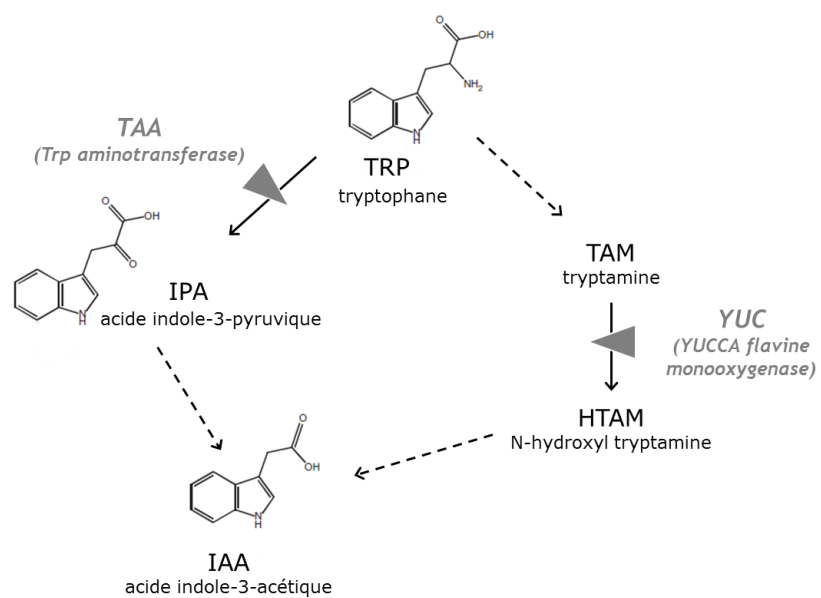


Figure II-11 – Schéma simplifié des voies de biosynthèse de l'auxine IAA.

MtGEA (Benedito et al., 2008; He et al., 2009). Toutefois, les transcrits de ces gènes sont principalement détectés dans l'albumen et/ou le tégument [tableau II-4]. La recherche d'homologues par "blast" de ces trois gènes identifie comme meilleur orthologue chez *A. thaliana* le gène At2g47140.1 codant une short-chain deshydrogenase reductase SDR5. Le blast sur les séquences de pois identifie le gène *SAD-C* (short-chain alcohol deshydrogenase) comme meilleur orthologue. Deux autres gènes *SAD-A* et *SAD-B* ont été identifiés chez le pois, avec un fort pourcentage d'homologie avec les SADs chez la tomate. D'abord identifiés comme ayant une expression induite dans les feuilles exposées aux radiations UV (Brosche and Strid, 1999), ces gènes ont également été décrits comme ayant une forte expression dans les graines (Scherbak et al., 2011). Le rôle de ces enzymes n'a pas été clairement identifié et leurs sites d'expression (tissus de la graine ou induite par les radiations UV dans les tissus exposés) et d'activité (présence de quinones ou de cycles aromatiques carbonés) semblent très restreints. Au vu du fort niveau de dérégulation observé dans les graines de mutant *dash* EMS109, on peut supposer que ces enzymes jouent un rôle important dans le développement de la graine. Ce rôle reste cependant à déterminer.

d - Gènes de la classe "35.not assigned"

Parmi les autres gènes validés par qPCR, trois (Medtr2g076980.1, Medtr8g046040.1 et Medtr6g009690.1) sont spécifiques de la graine d'après le Gene Atlas MtGEA (Benedito et al., 2008; He et al., 2009) et préférentiellement exprimés dans l'albumen. Leur expression étant réprimée chez le mutant *dash* EMS109, ces gènes pourraient être des cibles potentielles de *DASH*. Deux d'entre eux: Medtr2g076980.1 et Medtr8g046040.1 sont annotés comme des LCR related genes (low molecular weight cysteine rich). Ils codent des peptides courts d'environ 70 à 80 acides aminés et sont identifiés comme des DEFL (defensin like) (Tesfaye et al., 2013). Dans cette étude de Tesfaye et al. (2013), ces deux gènes sont également présentés comme spécifiques de la graine et non comme la plupart des DEFLs induits en présence de pathogènes dans les tissus inoculés ou dans les nodules lors d'associations symbiotiques. Il semblerait donc que, malgré les rôles importants que jouent les défensines dans les processus de défense face aux agressions par des pathogènes ou dans le développement des nodules, elles aient également un rôle dans les tissus reproducteurs et notamment dans le développement de la graine. Il est à noter la colocalisation des transcrits de la défensine Medtr8g046040 et de *DASH* au sein de la couche chalazale de l'albumen (cf figures II-1 et II-6, article), soulignant cette défensine comme cible possible de *DASH*.

II.3.3 - Recherche des gènes relatifs à la biosynthèse d'AIA

Le métabolisme hormonal est affecté chez le mutant *dash* EMS109 et l'une des hypothèses issue de l'analyse globale des données transcriptomiques est que cette dérégulation pourrait provenir de troubles soit au niveau du transport des auxines vers l'embryon, soit au niveau de la perception et de l'utilisation de cette hormone par les tissus de la graine. En complément, dans une approche ciblée, nous avons recherché

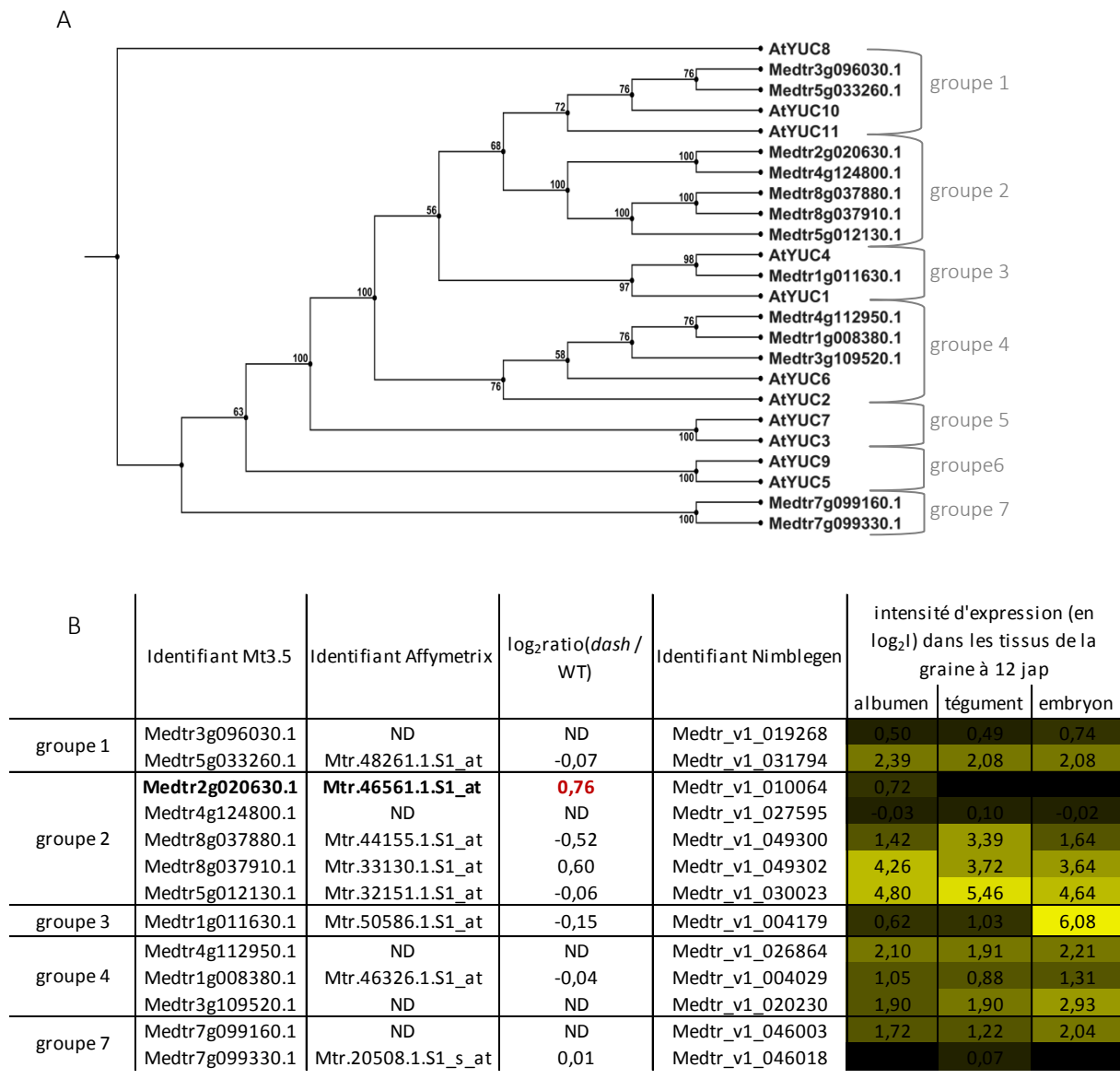


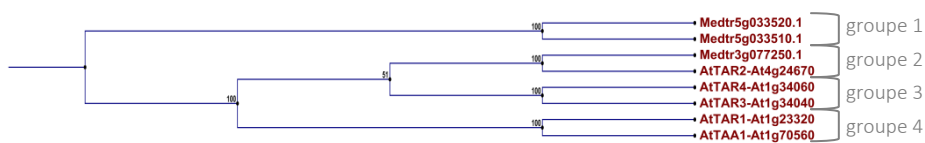
Figure II-12 – A. Arbre phylogénique des séquences protéiques YUCCA chez *A. thaliana* et *M. truncatula*.

B. Tableau présentant les niveaux d'expression des différents gènes YUCCA chez le mutant *dash* EMS109 par rapport au WT (log₂ratio) d'après les données Affymetrix de gousses à 8 jap (les valeurs p de probabilité pour l'ensemble de ces gènes sont supérieures à 0,05, excepté pour le gène Medtr2g020630.1 (p=0,0317)). Les niveaux d'expression dans les trois tissus de la graine à 12 jap (données NimbleGen – chapitre 3) sont indiqués pour chacun de ces gènes.

les gènes impliqués dans les voies de biosynthèse de l'AIA (acide indole acétique), ces gènes n'apparaissant pas dans la classe "17.hormone metabolism" mais pour la plupart dans la classe "26.7.misc.oxidases-copper, flavone, etc..". Un schéma simplifié de ces voies est présenté dans la figure II-11. Ce modèle, décrit jusqu'à présent dans de nombreuses études (LeClere et al., 2010; Tivendale et al., 2010) a été discuté, notamment quant aux rôles des flavines mono-oxygenases YUCCA qui interviennent dans la conversion des intermédiaires réactionnels (IPA en IAA ou TAM en HTAM) et d'autres modèles ont été proposés (Mashiguchi et al., 2011; Won et al., 2011) faisant intervenir les YUCCA dans plusieurs conversions de métabolites intermédiaires à la synthèse d'AIA.

Les gousses du mutant *dash* EMS109 présentent un niveau d'auxine important (cf. figure II-6, article), ce qui rappelle le phénotype de jeunes plantes sur-exprimant un gène *YUCCA* (AT4G32540.1 – *AtYUC1*) chez *A. thaliana*. Les auteurs ont en effet montré chez ces plantes une augmentation de la concentration en auxine (Zhao et al., 2001), suggérant l'implication des gènes *YUCCA* dans le contrôle de la teneur en AIA. De plus, ces gènes *YUCCA* (*YUC1*, *YUC4*, *YUC10* et *YUC11*) sont exprimés dans l'embryon pendant l'embryogénèse et seraient par conséquent liés à la régulation de l'AIA dans le développement de la graine (Cheng et al., 2007). Une recherche des gènes *YUCCA* a donc été réalisée chez *M. truncatula* en vue de définir des orthologues putatifs aux gènes *YUCCA* d'*A. thaliana* [figure II-12A]. Le gène *AtYUC1* est relativement proche de Medtr1g011630.1 en termes d'homologie de séquence (bootstrap de 97 – groupe 3). Les profils d'expression de ces deux gènes sont également comparables puisque le gène Medtr1g011630.1 est très fortement exprimé dans l'embryon à 12 jap (stade cotylédons linéaires) [figure II-12B] et le gène *AtYUC1* également, mais avec une intensité moindre, du stade globulaire au stade cotylédons linéaires. Afin d'étudier si les gènes *YUCCA* pourraient être à l'origine des fortes concentrations en auxines observées chez le mutant *dash* EMS109 (cf. figure II-6, article), nous avons analysé les niveaux d'expression de ces gènes dans l'étude comparative du transcriptome de gousses *dash* EMS109 vs WT (cf. tableau supplémentaire 1, article). Un seul (Medtr2g020630.1) apparaît dans cette liste des gènes significativement différenciellement exprimés avec une valeur p inférieure à 0,05 (cf analyse globale présentée dans l'article). Ce gène est sur-exprimé dans les gousses du mutant *dash* EMS109 mais est malheureusement très faiblement exprimé dans les tissus de la graine [figure II-12B]. Plusieurs de ces gènes montrent une expression dans les tissus de la graine, il n'est donc pas exclu que ces gènes présentent une dérégulation dans la graine qui ne soit pas significative en exploitant les données relatives aux gousses + graines. Ainsi, une baisse d'expression des gènes *YUCCA* dans la graine serait susceptible d'entraîner des troubles dans la distribution de l'auxine et de provoquer ainsi un avortement précoce des graines en début de développement des plantes *dash* EMS109 (pendant la phase où l'ensemble des graines, et par conséquent les gousses avortent). Les gènes *YUCCA* ayant des fonctions redondantes (Zhao et al., 2001), on peut supposer que plusieurs autres gènes de cette famille assurent les fonctions des gènes *YUCCA* dérégulés en prenant "le relais", conduisant ainsi à une production de graines (et de gousses) plus

A



B

	Identifiant Mt3.5	Identifiant Affymetrix	log ₂ ratio(<i>dash</i> / WT)	Identifiant Nimblegen	intensité d'expression (en log ₂) dans les tissus de la graine à 12 jap		
					albumen	tégument	embryon
groupe 1	Medtr5g033520.1	Mtr.62.1.S1_at	0,93	Medtr_v1_031818	0,54		0,39
	Medtr5g033510.1	Mtr.25150.1.S1_at	-0,42	Medtr_v1_031817	7,22	3,87	2,72
groupe 2	Medtr3g077250.1	Mtr.17173.1.S1_s_at	-0,94	Medtr_v1_017978	1,14	1,23	5,18

Figure II-13 – A. Arbre phylogénique des séquences protéiques TAA/TAR chez *A. thaliana* et *M. truncatula*.

B. Tableau présentant les niveaux d'expression des différents gènes TAA/TAR chez le mutant *dash* EMS109 par rapport au WT (log₂ratio) d'après les données Affymetrix de gousses à 8 jap (les valeurs p de probabilité pour l'ensemble de ces gènes sont supérieures à 0,05). Les niveaux d'expression dans les trois tissus de la graine à 12 jap (données NimbleGen – chapitre 3) sont indiqués pour chacun de ces gènes.

tardivement. On pourrait également envisager que les fonctions de l'ensemble des gènes YUCCA n'étant pas toutes rétablies, il subsiste un retard de développement de l'embryon (cf. figure II-5 – article) qui induit la production de petites graines.

Concernant les gènes de la famille TAA/TAR (tryptophane aminotransferase of *arabidopsis*/ TAA related protein), également impliqués dans les voies de biosynthèse de l'auxine [figure II-11], trois gènes ont été identifiés chez *M. truncatula* et la phylogénie avec les séquences d'*A.thaliana* montre clairement que Medtr3g077250.1 est un orthologue de TAR2 [figure II-13A]. Les triple mutants pour les gènes TAA1, TAR1 et TAR2 (mutants d'insertion T-DNA) montrent des défauts d'accumulation d'AIA accompagnés d'anomalies dans le développement de l'embryon (Stepanova et al., 2008). Deux des trois gènes TAR chez *M. truncatula* montrent une expression préférentielle soit dans l'albumen (Medtr5g033510.1) soit dans l'embryon (Medtr3g077250.1) à 12 jap [figure II-13B], suggérant une possible implication dans le développement de la graine. Ces gènes ne montrent pas de dérégulation dans les gousses *dash* EMS109 à 8 jap [figure II-13B], mais comme mentionné précédemment, il se pourrait que des dérégulations aient lieu dans les tissus de la graine.

II.3.4 - Applications d'auxines exogènes

En vue de réverser le phénotype des mutants *dash* EMS109, l'application exogène de plusieurs auxines a été testée. Les traitements ont été appliqués sur les gousses en cours de développement à 6 et 8 jap, par immersion quelques secondes dans une solution contenant les hormones.

L'hormone synthétique 2-4D a été utilisée à une concentration de 100 mg.L^{-1} et les mesures de poids d'une gousse et d'une graine relatives à des gousses traitées (2-4D+) et non traitées (2-4D-) d'une même plante ont été comparées (récoltes réalisées tout au long du cycle, [figure II-14]). Des doses de 2-4D trop importantes chez le WT auraient un effet toxique sur le développement de la gousse et de la graine (réduction du poids d'une graine et d'une gousse) alors qu'un apport passif de 2-4D (*i.e.* 2-4D-, mesures sur les gousses non traitées) aurait peu d'effet. Concernant le mutant *dash* EMS109, la dose reçue par les gousses traitées directement (*i.e.* 2-4D+) semble être trop forte. En effet, bien que le poids d'une gousse ne varie pas, ces gousses ne contiennent que très peu de graines qui apparaissent nécrosées (noires et sèches). L'observation des mesures faites sur les gousses du mutant *dash* EMS109 non directement traitées (*i.e.* 2-4D-) suggère une restauration partielle du phénotype. En effet, sans atteindre les valeurs observées chez les plantes sauvages, les graines ont un poids relativement similaire à celui des graines obtenues chez le mutant *dash* EMS109 en fin de cycle. Il semblerait également que le nombre de graines à l'intérieur des gousses soit supérieur. Il est à noter que les mesures de poids de gousses, nombre de graines par gousse et poids de graines pour les gousses non directement traitées proviennent d'un pool de gousses recueillies tout au long de cette expérimentation. Afin d'être plus rigoureux et de pouvoir interpréter plus précisément ces résultats, il aurait fallu récolter les gousses dont le développement est synchronisé, c'est-

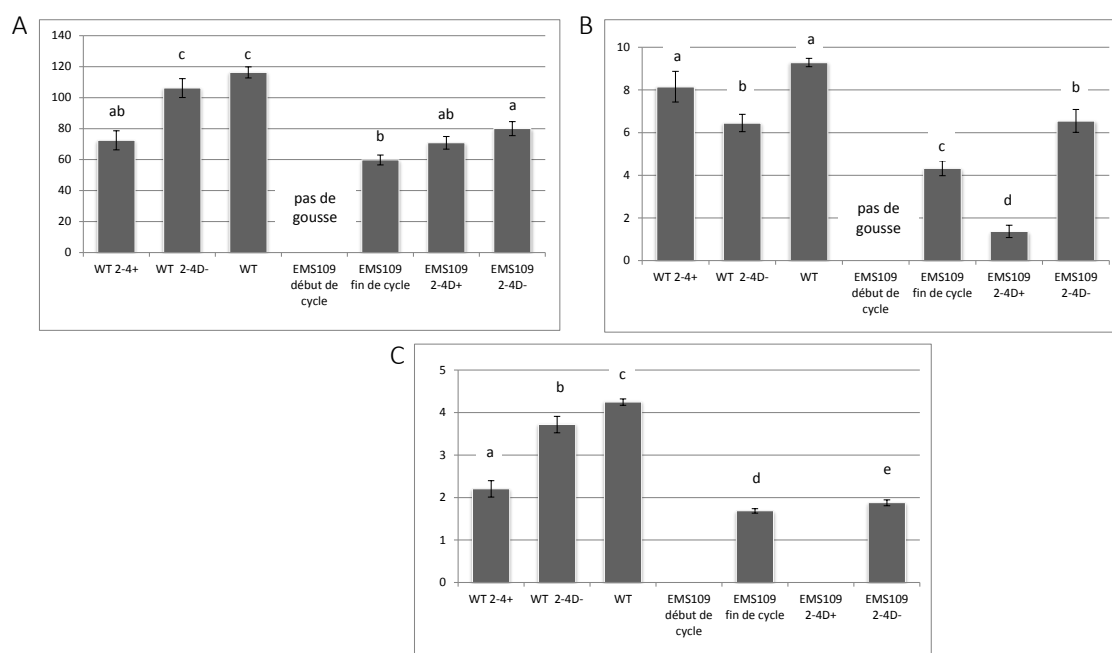


Figure II-14 – Influence du traitement 2-4D (100mg.L^{-1}) sur le développement des gousses et des graines. (A) Mesure du poids d'une gousse en mg. (B) Nombre de graine par gousses. (C) Poids d'une graine en mg. 2-4D+ fait référence aux mesures faites sur des gousses directement traitées et 2-4D- concerne les autres gousses se trouvant sur la même plante. Les récoltes des gousses pesées ont été réalisées tout au long du cycle.

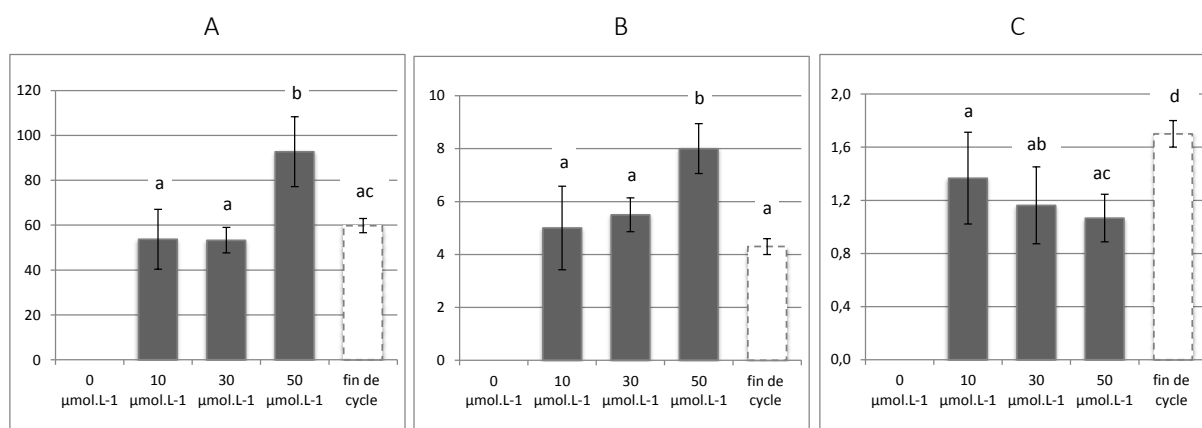


Figure II-15 – Influence du traitement AIA sur le développement des gousses et des graines du mutant *dash* EMS109. (A) Mesure du poids d'une gousse en mg. (B) Nombre de graine par gousses. (C) Poids d'une graine en mg. Les gousses ont été traitées avec différentes concentrations d'AIA.

à-dire séparer, après traitement indirect de 2-4D, les gousses obtenues en cours de développement (pendant que les gousses des mutants *dash* EMS109 non traitées avortent) des gousses obtenues en fin de cycle. Les résultats obtenus ne permettent pas de conclure quant à l'efficacité du traitement, mais seulement de proposer une hypothèse quant à la toxicité des doses utilisées car les gousses non directement traitées semblent mieux "réagir" au traitement (diminution du poids d'une gousse et d'une graine pour les gousses WT traitées et nécrose des graines provenant des gousses EMS109 traitées). Il pourrait donc y avoir un effet systémique des hormones, qui conduirait à favoriser le développement des graines ayant reçu peu d'auxine par rapport à des graines contenues dans des gousses directement traitées. Il est à noter que les plantes produisent des gousses viables plus tôt au cours de leur cycle de développement. Le mécanisme qui se met habituellement en place après quelques semaines chez les mutants *dash* EMS109 afin d'échapper à l'avortement des gousses est donc restauré plus tôt grâce à l'application exogène de l'auxine.

Une autre expérimentation a été menée en vue de tester une hormone naturelle: l'AIA. Les résultats sont présentés dans la figure II-7 de l'article. Aucune différence significative dans le poids d'une gousse et d'une graine n'a été observée entre les mesures faites sur les gousses directement traitées et les gousses qui l'ont été indirectement, les résultats présentés étant ceux des gousses directement traitées. Comme précédemment pour 2-4D, les plantes dont certaines gousses ont été traitées produisent plus tôt des gousses viables mais nous ne pouvons conclure de ces résultats qu'à une restauration partielle par l'application de l'AIA. En effet, le poids de ces gousses et des graines qu'elles contiennent reste inchangé par rapport aux poids observés pour les gousses et graines produites en fin de cycle chez le mutant *dash* EMS109. Les mêmes interrogations persistent quant à la dose d'hormone utilisée, c'est pourquoi de nouveaux tests ont été réalisés avec des applications d'AIA à des concentrations inférieures, de 10 à 50 $\mu\text{mol.L}^{-1}$ (l'AIA dilué à 100 mg.L^{-1} équivaut à une concentration de 571 $\mu\text{mol.L}^{-1}$) [figure II-15]. Ces résultats préliminaires sont à interpréter avec réserve car ces mesures concernent seulement les pesées de 5 à 10 gousses (et leur contenu en graines) par traitement. Il faudrait donc renouveler ces tests afin de s'assurer des résultats et de réduire les écart-types. Ces graphiques suggèrent tout de même que l'on arrive à réverser le phénotype (dans le sens de rendre plus précoce la production de gousses viables chez le mutant) même avec des concentrations jusqu'à dix fois inférieures à celles utilisées précédemment. Il semblerait que des applications à une concentration de 50 $\mu\text{mol.L}^{-1}$ soient suffisantes pour augmenter le poids d'une gousse par rapport aux poids d'une gousse produite en fin de cycle. De façon surprenante, ce poids d'une gousse plus élevé serait dû à un nombre de graines par gousse plus important (le poids d'une graine étant amoindri) et/ou à une augmentation du poids du tissu constituant la gousse. Les troubles au niveau du développement de la graine observés chez le mutant pourraient être dus à des troubles de la perception de l'auxine par les gousses et/ou à un défaut de communication entre gousses et graines qui

feraient que le poids d'une gousse augmente au dépens du poids des graines de cette gousse. Du fait que cette expérience n'ait pas été renouvelée, ces informations restent à valider.

Les traitements avec un inhibiteur du transport d'auxine, le TIBA sur des gousses de plantes WT ont montré une diminution du poids d'une gousse et du poids d'une graine mais pas de variation du nombre de graines par gousse contrairement aux observations précédentes chez la tomate (Serrani et al., 2010). Cet inhibiteur de transport d'efflux des auxines conduit à une augmentation de la concentration en AIA dans les fruits et à une réduction de la taille des graines et des fruits (Hamamoto et al., 1998), phénotype similaire à celui observé chez les mutants *dash* EMS109 (cf figure II-7A, article) où le niveau d'AIA est déjà très élevé dans les gousses. Ceci suggère que la mutation *DASH* engendre entre autre des troubles du transport d'auxine proche de ceux observés dans le cas d'une inhibition du transport d'auxine, mais à moindre échelle (le phénotype des gousses et graines WT traitées au TIBA n'étant pas similaire à celui observé chez les mutants). Le TIBA induit également des troubles au niveau de l'embryogénèse conduisant au phénotype cotylédons fusionnés (Liu et al., 1993 ; Choi et al., 1997) observés chez le mutant *dash* EMS109. Ce phénotype est similaire à des mutations des transporteurs d'auxine PIN. Parmi les dix transporteurs PIN identifiés chez *M. truncatula* (Schnabel and Frugoli, 2004), quatre sont retrouvés dans l'analyse comparative du transcriptome comme significativement ($p < 0,05$) sous-exprimés chez le mutant *dash* EMS109 [annexe II-6]. Ces transporteurs ne montrent cependant pas de forte dérégulation (de 0,5 à 2,5 fois moins exprimés chez le mutant par rapport au type sauvage). Bien que ces transporteurs ne montrent pas d'expression préférentielle dans la graine, les données issues des analyses NimbleGen (cf chapitre III) indiquent que deux d'entre eux (MtPIN3-AY115838.1 et MtPIN4-Medtr6g069510.1) ont une expression dans la graine préférentielle à l'embryon, ce qui suggère un possible dysfonctionnement dans le transport d'AIA vers l'embryon. Afin de valider cette hypothèse il serait intéressant de confirmer les variations des profils d'expression de ces transporteurs dans des échantillons de tissus de la graine isolés, la perte d'intensité du signal étant probable dans l'interprétation des données sur des échantillons de gousses entières.

II.4 - Matériels et méthodes

II.4.1 - Détermination de la teneur en sucre

La teneur en sucres solubles (saccharose, fructose et glucose) a été mesurée dans les graines de mutant *dash* (EMS 109 et NF6042) et leurs WT respectifs (issus des génotypes A17 et R108). Les plantes ont été cultivées sous serre dans des pots de 1,5L rempli d'un mélange attapulgit : billes d'argiles (60 :40). Les conditions de température et de luminosité étaient régulées : 20°C le jour avec une photopériode de 16 heures et 18°C la nuit. L'apport en azote n'était pas limitant tout au long du développement des plantes, avec un apport 3,5N/3,1P/8,6K. Les fleurs ont été notées et les gousses collectées aux stades 6, 8, 10 et 12

jap selon Gallardo et al. (Gallardo et al., 2007). Les graines ont été disséquées sous loupe binoculaire dans des boîtes de Pétri placées sur de la glace, et les différents échantillons ont été pesés puis congelés dans l'azote liquide avant d'être stockés à -80°C. Au total, cinq réplicats biologiques pour chaque stade et pour chaque génotype étudié ont été générés. Ces échantillons de graines ont ensuite été broyés en tube Eppendorf 2,0mL avec une bille de verre grâce à un broyeur de type Retsh puis lyophilisés pendant 24 h. Afin de déterminer de façon précise la teneur en sucres dans des échantillons n'ayant pas tous la même masse (entre 1 et 17 mg), 20 µL d'une solution de mélézitose à 4 g/L, servant de témoin interne sont ajoutés. 500µL d'une solution de méthanol 80 % sont également ajoutés au broyat afin de sécher les échantillons, qui sont ensuite incubés 15 min à 76°C, puis le méthanol est évaporé au speed-vac (Mi Vac quatre concentrator, GENEVAC) pendant 2 h à 40°C. Le culot est repris dans 150µL d'eau ultra-pure et centrifugé 5 min à 4°C à 13500 g (R5417, Eppendorf), le surnageant est ensuite dilué au ¼ dans un volume final de 200µL. Les échantillons sont analysés par chromatographie ionique haute performance en phase liquide, en utilisant le système DIONEX (Dionex Corp., Sunnyvale, CA, USA), contenant une précolonne AS11 ion exchanger et une colonne Dionex CarboPac PA1 (2 x 250 mm). Les sucres sont détectés par ampérométrie après séparation sur une colonne CarboPac PA1 avec un débit constant à 0.25mL/min de NaOH 0,1 M. La concentration en sucres des échantillons est déterminée par comparaison avec un échantillon standard dont les quantités de sucres sont connues. Les teneurs en sucres sont ensuite exprimées en mg de sucre par mg de poudre de graine.

Les dosages de sucres ont été réalisés au laboratoire IRHS à Angers sous la responsabilité de Julia BUITINK.

II.4.2 - Application d'auxine exogène

Les conditions de culture des plantes sont identiques à celles précédemment décrites. Les fleurs ont été notées en tout début de floraison et les gousses en cours développement ont été traitées. Les traitements consistent en une immersion quelques secondes dans une solution contenant les auxines AIA et 2-4D (différentes concentrations utilisées, cf figure II-14 et II-15) afin de favoriser la pénétration de l'auxine, une goutte de Tween est ajoutée pour 50mL de solution. Chaque traitement a été répété deux fois à 6 et 8 jap. Les gousses ont été récoltées à maturité et les mesures de poids de gousses, nombre de graines par gousse et poids de graines ont été déterminées (environ 30 gousses et leurs contenus en graines pour le traitement 2-4D, et seulement 5 à 10 gousses pour les traitements AIA).

II.5 - Conclusion

L'expression préférentielle de *DASH* dans la graine et plus particulièrement dans l'albumen est associée à une spécificité d'expression au stade embryogénèse, phase importante pour le développement de l'embryon et la détermination de la taille de la graine [figure II-1]. La caractérisation de ce gène *DASH* a

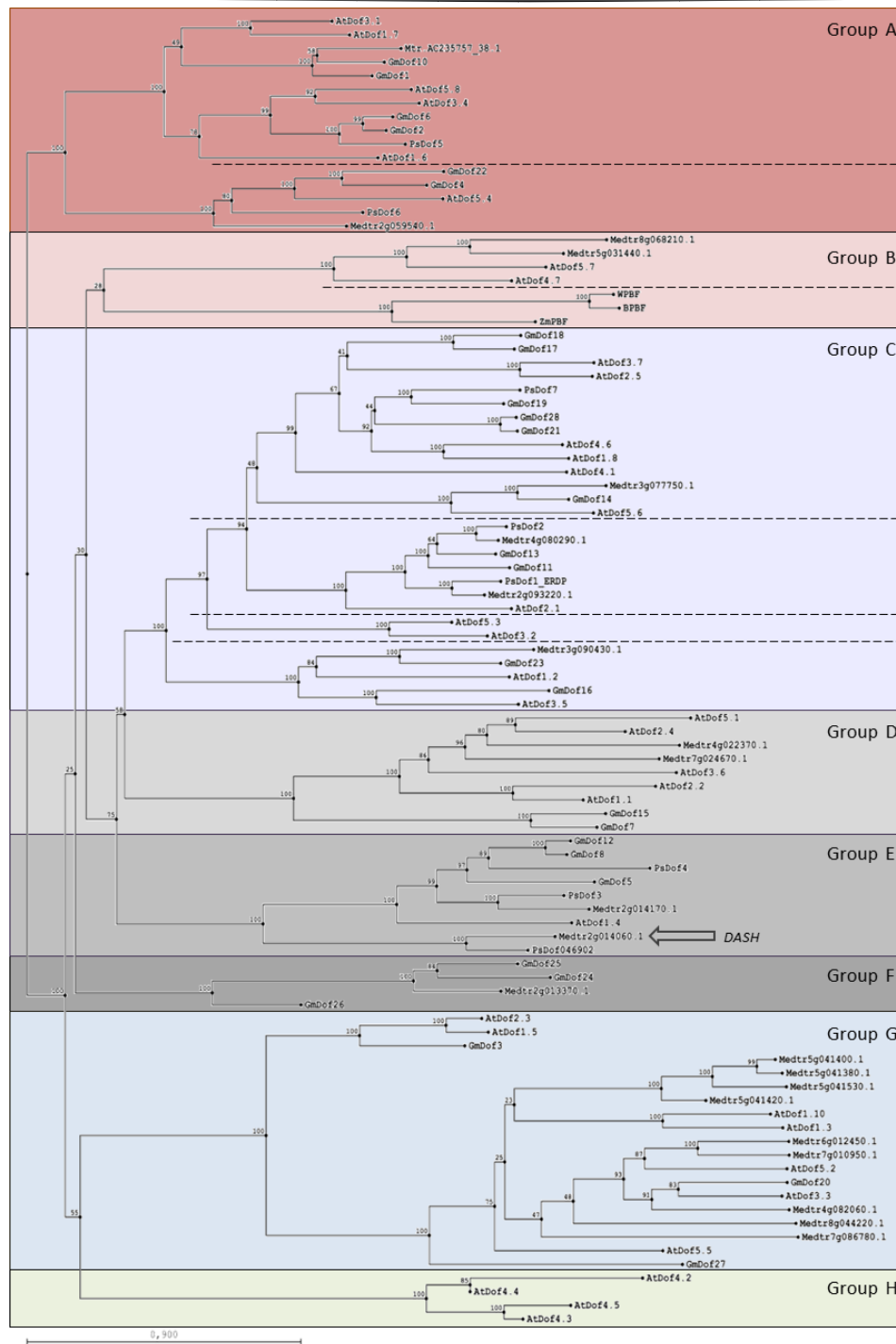
été entreprise à l'aide de deux mutants *dash* (EMS109 et NF6042) présentant des phénotypes forts. Chez le mutant *dash* EMS109 homozygote pour la mutation, le développement des gousses et graines est bloqué très tôt une fois les fleurs fanées pendant les premières semaines de vie de la plante, puis progressivement des gousses et graines parviennent à se développer jusqu'au stade 8-10 jap avant d'avorter également. Enfin, un processus d'échappement qui reste encore à éclaircir permet un développement normal des gousses et graines jusqu'au stade mature, avec une production de graines plus petites que les graines de type sauvage [figure II-2III]. Quant au mutant *dash* NF6042, seules des plantes hétérozygotes ont pu être générées, la descendance de ces plantes conduisant exclusivement à des plantes hétérozygotes et sauvage dans des proportions 1:3 [annexe II-2III], ce qui laisse supposer que la mutation à l'état homozygote est létale. Cette hypothèse a été confortée par le comptage du nombre de petites graines dans des gousses hétérozygotes en cours de développement conduisant à une ségrégation 1:3 [annexe II-2I]. L'étude cytologique des différents stades de développement associés à l'embryogénèse (de 6 à 12 jap) a également démontré un fort impact de la mutation sur le développement de l'embryon, d'abord retardé par rapport au WT (8 jap) puis complètement stoppé (10 jap) pour l'ensemble des graines produites en début de cycle de développement pour le mutant *dash* EMS109 [figure II-5] et pour les petites graines homozygotes contenues dans les gousses hétérozygotes du mutant *dash* NF6042, les autres graines dans ces mêmes gousses se développant dans des conditions comparables au sauvage [annexe II-4]. Les observations des graines produites en fin de cycle du mutant *dash* EMS109 montrent un développement comparable à celui des graines sauvage, mais avec un phénotype cotylédons fusionnés [figure II-5], que l'on retrouve également lors des semis [figure II-3]. Ce phénotype cotylédons fusionnés rappelle celui observé chez des mutants *cuc* et soulève la question de l'importance des hormones dans le développement de l'embryon (Aida et al., 1997; Moller and Weijers, 2009). L'étude comparative du transcriptome des gousses du mutant *dash* EMS109 par rapport au WT à 8 jap a également démontré une dérégulation importante du métabolisme hormonal [tableau II-2, figure II-9 et annexe II-8]. Plusieurs hypothèses sont envisageables quant aux causes possibles des dérégulations observées. Le métabolisme des gènes de réponse aux auxines est fortement sous-exprimé chez le mutant *dash* EMS109 mais ceci n'est pas dû à un manque d'auxine car les dosages d'hormone ont révélé au contraire une forte teneur en auxine dans les gousses [figure II-7]. On peut alors penser qu'il s'agit d'un trouble de la perception de l'auxine par les tissus de la graine ou au niveau du transport des auxines de la gousse vers la graine et plus précisément vers l'embryon (dérégulation de certains gènes PINs - tableau supplémentaire 1 article) si l'on considère que les auxines sont concentrées dans la gousse. D'après des données de la littérature indiquant que les graines constituent également une source d'auxine pour le développement du fruit (Sundberg and Ostergaard, 2009), on pourrait également envisager que ces fortes teneurs en auxine sont confinées dans la graine, empêchant ainsi le développement de la gousse. Afin de pouvoir interpréter plus justement ces données, il faudrait doser séparément l'auxine présente dans les graines et l'auxine présente dans le tissu "gousse"

afin de déterminer si le problème provient du transport ou de la perception. On observe également une forte teneur en auxine conjuguée AIA-Asp dans les graines du mutant *dash* EMS109 par rapport au WT. Ces conjugués sont des formes d'auxine inutilisables par la plante et sont destinés à être dégradés (Ludwig-Muller, 2011), ce qui conforte l'hypothèse que les auxines ne sont pas utilisées par les tissus de la graine pour leur développement mais sont directement dégradées. L'auxine entraîne la liaison des protéines Aux/IAA au complexe SCF^{TIR1} qui intervient dans la dégradation des protéines marquées par l'ubiquitine (Dharmasiri et al., 2005; Kepinski and Leyser, 2005), cette voie de dégradation, associée au code BIN "29.5.11.protein.degradation.ubiquitin", ressort d'ailleurs significativement fortement surexprimée dans les gousses du mutant *dash* EMS109 suite à l'analyse transcriptomique [tableau II-2 et figure II-9]. Le lien entre le métabolisme hormonal et le gène *DASH* n'est pas encore clairement défini mais de nombreuses pistes suggèrent que ce FT intervient dans les voies métaboliques de transport ou de perception de l'auxine. Il reste encore de nombreux points à éclaircir afin de mieux comprendre le rôle de *DASH* dans le développement de la graine et de l'embryon et de pouvoir ainsi interpréter les fonctions de ce gène dans la réponse à l'auxine.

Il serait intéressant d'observer les semis de graines *dash* EMS109 traitées par AIA ou 2-4D afin de noter si le phénotype cotylédons fusionnés persiste dans ces graines et si l'on obtient un pourcentage de dans le développement des plantules rétabli (40% de phénotype cotylédons fusionnés chez les graines du mutant *dash* EMS109 et 90% de ces graines ne présentent pas d'apex). De même il semblerait judicieux de vérifier que le métabolisme hormonal est affecté chez le mutant hétérozygote NF6042 et de réaliser des tests d'application exogènes d'hormones afin d'observer si le nombre de graines par gousse augmente et par conséquent si l'application d'auxine parvient à "sauver" les petites graines.

Plus largement, il serait intéressant d'observer si d'autres critères de qualité de graine comme la teneur en protéine, teneur en sucres ou en lipides sont affectés chez les mutants. Des expérimentations afin de déterminer la composition protéique ont été conduites dans ce but au laboratoire. Les premiers résultats n'indiquent pas de changements dans la composition en protéine de réserve (de type vicilline ou légumine) mais la composition en protéines de maturation telles que les LEA semble affectée puisque l'on retrouve une moindre accumulation de ces protéines dans les graines matures de mutant *dash* EMS 109. D'autres protéines telles que des sucrose-binding protein sont également moins présentes, les dosages de certains sucres dans les graines ayant déjà permis de détecter ce défaut d'accumulation de glucose, fructose et saccharose. L'étude du protéome des graines serait donc une piste intéressante à poursuivre afin d'analyser plus finement les défauts d'accumulation dans les graines de mutant *dash* (EMS109 et NF6042).

II.6 - Annexes



Annexe II-1 - Dof proteins relationships. The phylogenetic tree was created using full-length Dof coding sequences: 36 from *Arabidopsis thaliana* (At), 27 from *Glycine max* (Gm), 22 from *Medicago truncatula* (Medtr), 8 from *Pisum sativum* (Ps) and PBF sequences from cereals: *Zea mays* (ZmPBF), *Hordeum vulgare* (BPBF) and *Triticum aestivum* (WPBF). These 96 protein sequences were aligned and organized in a tree based on sequence similarities from the ClustalW alignments of the full-length DOF protein sequences generated with CLC genomics workbench (gap open cost:10 and gap extension cost:1) (CLC bio -CLC sequence viewer 6.2, Aarhus, Denmark).

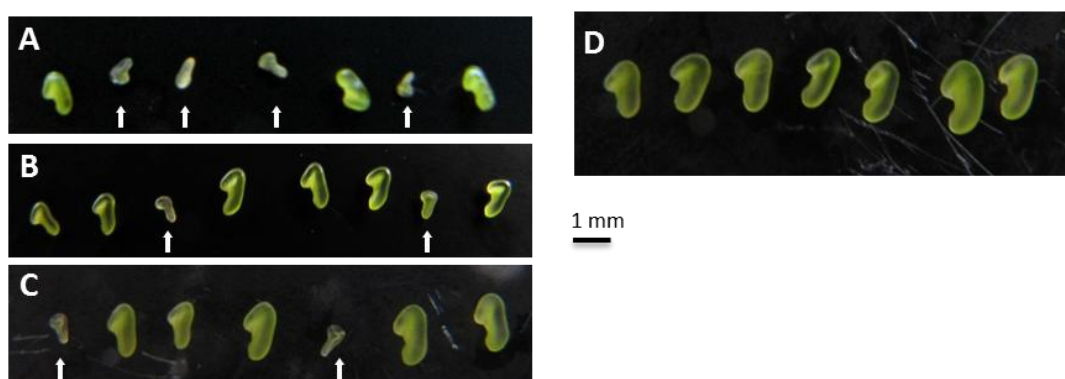
I

Mutant dash / + segregation		homozygous dash / dash	heterozygous dash / +	wild-type + / +	total	χ^2 (2 : 1)	P-value
EMS109	observed values	1	66	34	101	0,0297	0,8632
	predicted values	0	67	34	101		
NF6042	observed values	0	79	45	124	0,4879	0,4849
	predicted values	0	83	41	124		

II

Mutant	Genotype	Frequency of small seeds (%)	Observed / total	χ^2 p-value for 3:1
NF6042	+/+	4,1	7/172	2,31E-10
	+/-	28,4	105/370	0,133
EMS 109	+/+	0,5	1/181	2E-14
	+/-	27,9	33/118	0,7969

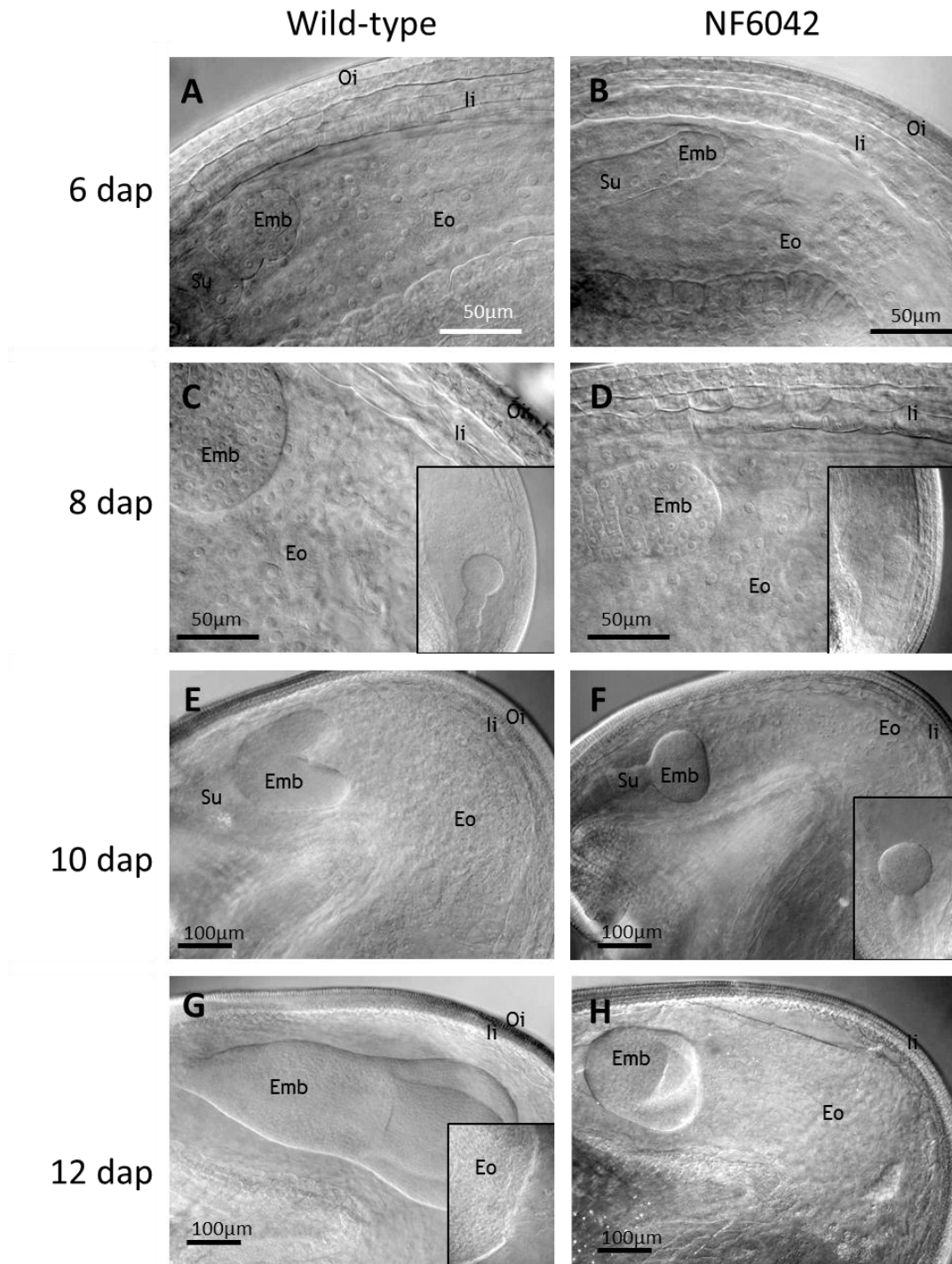
III



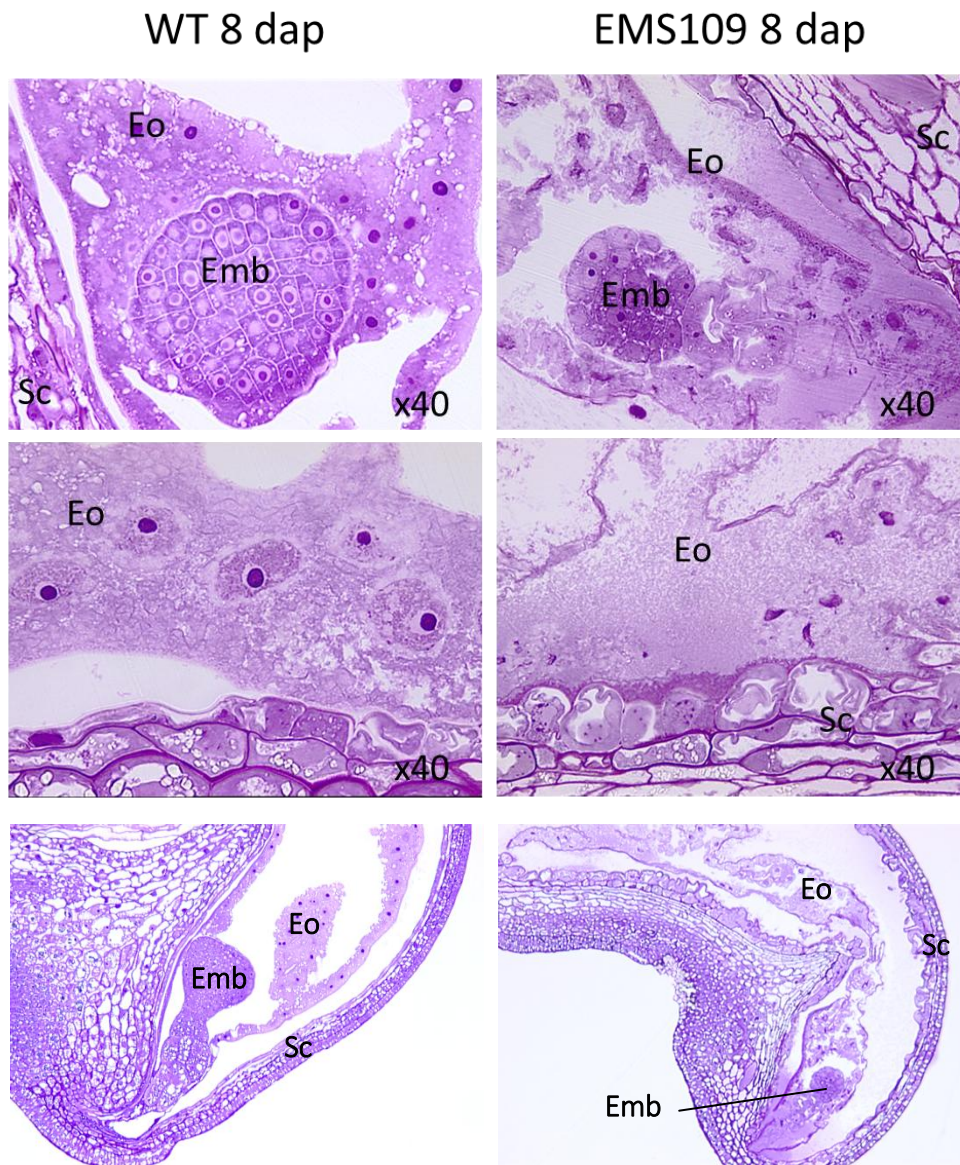
Annexe II-2 - (I) Segregation observed for EMS109 and NF6042.

(II) Table with seed phenotype segregation in pods of heterozygous NF6042 and EMS109. Normal and small seed distribution are compared to a theoretical χ^2 distribution of 3:1.

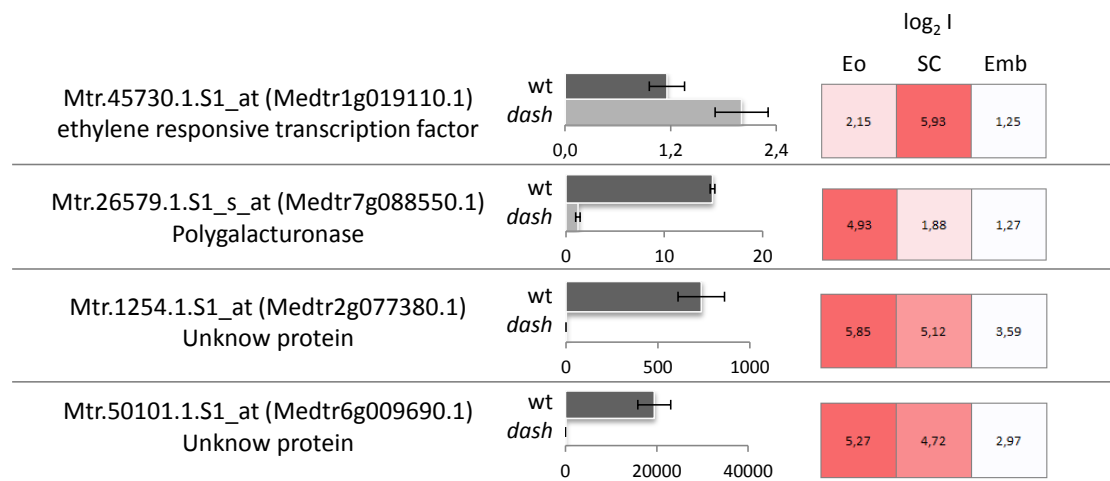
(III) Phenotype of NF6042 seeds dissected from NF6042 heterozygous pods at 6 dap (A, B), 8 dap (C), from WT at 8 dap (D). Arrows indicate smaller seeds corresponding to putative homozygous seeds.



Annexe II-3 - Phenotypical characterization of the NF6042 seed phenotype. Cleared pictures of embryo and endosperm through seed development. WT: (A, C, E, G) and heterozygous NF6042 mutant (B, D, F, H). Globular stage corresponding to 6 dap (A, B). Transition stage, corresponding to 8 dap (C, D). Heart stage, corresponding to 10 dap (E, F). Linear cotyledon stage corresponding to 12 dap (G, H). Emb, embryo; Eo, endosperm; li, internal integument; Oi, outer integument; Su, suspensor. Scale bar indicated on pictures.



Annexe II-4 - Microscopic observations of 8 DAP seeds from mutant *dash* (EMS109) and WT.



Annexe II-5 - qRT-PCR expression profile of genes deregulated in 8 dap seeds of *dash* (EMS109) and WT. Expression in seed tissue (unpublished data). Eo:endosperm, SC:seed coat, Emb:embryo.

BIN code	BIN Name	id affymetrix	gene_id_v3.5	log2(rpkm (data/ctrl))	annotation	Leaf	Petal e	Vegulu e	Stem	Flower	Pod	Seed_1day	Seed_2day	Seed_3day	Seed_4day	Seed_5day	Root	Seed_C	Endo	Sc	Er		
17.1.1	hormone metabolism,abscisic acid synthesis-degradation	Met5486.151_x_at	medtrd9028930.1	10	HVA2.1-like protein I	126	282	899	697	119	863	199	192	314	244	222	144	158	178	4.4	4.2	6.0	
		Met2957.151_x_at	medtrd9028930.1	10	Aldehyde oxidase	108	110	72	71	77	77	88	113	109	112	180	260	427	80	204	4.2	4.2	6.0
		Met4006.151_x_at	medtrd9044601.1	1.1	Malic dehydrogenase-like protein sulfatase	110	110	75	39	79	131	53	69	60	56	39	82	181	45	4.4	4.4	6.0	
		Met539.151_x_at	medtrd9039010.1	1.2	Maldehyde oxidase	11	30	23	16	37	25	57	37	40	35	33	93	55	41	4.4	4.4	6.0	
		Met3891.151_x_at	medtrd9058010.1	1.3	Maldehyde oxidase-like protein sulfatase	110	110	75	39	79	131	53	69	60	56	39	82	181	45	4.4	4.4	6.0	
17.1.1.1	hormone metabolism,abscisic acid synthesis-degradation,synthesis zeaxanthin epoxidase	Met4995.151_x_at	AC146630.4.1	1.5	HVA2.1-like protein I	86	87	92	85	115	96	268	264	378	666	1033	500	600	759	6.1	3.8	3.8	
		Met4064.151_x_at	AC146630.4.1	1.5	HVA2.1-like protein I	81	55	70	54	83	49	206	179	332	580	936	1097	598	625	4.2	4.2	6.0	
		CUTS1752_10202.1	NA	0.7	Zeaxanthin epoxidase	100	100	209	96	20	21	28	28	28	28	28	28	28	28	4.2	4.2	6.0	
		Met32679.151_x_at	NA	0.8	Zeaxanthin epoxidase	4302	2326	850	893	327	450	150	229	427	389	321	139	79	139	79	139	79	
		Met190.151_x_at	medtrd9022800.1	0.8	Zeaxanthin epoxidase	15	19	16	20	102	43	104	129	196	145	98	22	67	332	4.2	7.0	7.0	
17.1.2	hormone metabolism,abscisic acid signal transduction	Met4106.151_x_at	NA	1.7	ABA8	4462	1249	1119	479	238	232	365	287	295	277	229	279	279	279	279	279	279	
		Met4588.151_x_at	medtrd1017801.1	1.6	ABSCISIC ACID-INSENSITIVE 5-like protein	417	318	318	315	316	329	417	418	366	286	285	104	297	871	6.1	6.2	6.2	
		Met22184.151_x_at	medtrd9082301.1	0.8	ABSCISIC ACID-INSENSITIVE 5-like protein	220	301	281	276	468	426	564	544	518	517	475	367	317	926	8.9	6.5	6.5	
		Met4585.151_x_at	NA	1.4	ABA8	213	175	32	125	288	31	43	28	24	20	27	58	110	43	4.2	4.2	6.0	
		Met28794.151_x_at	NA	1.6	ABA8	269	189	40	149	491	30	70	47	33	31	34	72	161	49	2.1	2.1	6.0	
17.2.1	hormone metabolism,auxin synthesis-degradation	Met27439.151_x_at	medtrd1050601.0	0.9	IAA-amino acid hydrolase IAR1-like protein	47	71	66	78	34	1093	5873	402	3297	1677	974	132	221	4201	2.1	2.1	6.0	
		Met45439.151_x_at	AY115838.1	-1.4	Auxin efflux carrier protein - PMPN3	1444	978	510	630	226	89	18	45	60	81	35	16	27	21	2.1	2.1	6.0	
		Met46539.151_x_at	medtrd9039701.1	-1.1	Auxin efflux carrier protein - PMPN3	20	28	45	45	67	92	17	34	15	16	15	14	15	16	1.8	1.8	6.0	
		Met1076.151_x_at	medtrd9058010.1	-1.2	Auxin efflux carrier protein - PMPN3	1563	1013	305	149	215	1564	268	268	268	268	268	268	268	268	268	268	268	
		Met3874.151_x_at	AY115838.1	-1.6	Auxin efflux carrier protein - PMPN3	882	748	595	600	367	425	605	602	391	186	190	114	283	40	114	283	40	114
17.2.2	hormone metabolism,auxin signal transduction	Met2561.151_x_at	NA	-0.7	NA	14	15	29	14	9	17	12	14	14	13	12	14	24	4	5.1	5.1	6.0	
		Met237.151_x_at	medtrd9042201.1	-0.8	NOD26 acting enzyme E1 regulatory subunit	100	100	209	96	20	21	28	28	28	28	28	28	28	28	28	4.2	4.2	6.0
		Met2109.151_x_at	medtrd9012901.1	0.8	Protein AUSEIN SIGNALING F-BOX	403	1027	1084	1002	1438	636	236	369	455	412	480	575	620	563	6.1	6.1	6.0	
		Met4808.151_x_at	NA	0.9	NA	1253	1292	1070	1728	1778	1738	1738	1738	1738	1738	1738	1738	1738	1738	1738	1738	1738	
		Met1340.151_x_at	medtrd9022001.1	1.4	Auxin influx protein	219	1137	453	990	3762	403	314	561	444	401	293	80	1054	944	8.8	8.8	8.8	
17.2.3	hormone metabolism,auxin-induced-regulated-responsive-activated	Met19818.151_x_at	medtrd9072401.1	1.8	Auxin-induced protein 6B	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	
		Met49768.151_x_at	medtrd9072401.1	1.8	Auxin-induced protein 6B	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	
		Met49793.151_x_at	medtrd9072401.1	1.8	Auxin-induced protein 6B	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	
		Met49793.151_x_at	medtrd9072401.1	1.8	Auxin-induced protein 6B	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	
		Met49793.151_x_at	medtrd9072401.1	1.8	Auxin-induced protein 6B	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	120	

Annexe II-6 - List of differentially expressed genes belonging to deregulated hormone functional classes. Log2 ratio of mutant data versus WT data from three repetitions of Affymetrix GeneChip is presented followed by data downloaded from Medicago GeneAtlas (<http://mtgea.noble.org/v3/experiments> from Benedito et al., 2008 and seed coat data from Pang et al., 2008. 12 dap seed tissue expression data (tegument, endosperm and embryo) from a Nimblegen array are indicated. endo: endosperm, sc:seed coat, emb: embryo

BinCode-BinName	pourcentage de répartition des sondes par classe pour l'ensemble de la puce	analyse sondes significativement différentiellement exprimées chez le mutant <i>dash</i> (EMS109) par rapport au WT p<0,01		estimation de la dérégulation d'une classe (enrichissement ou appauvrissement en gènes)	valeur p associée à la dérégulation de la classe
		nombre de sondes par classe fonctionnelle	pourcentage de répartition des sondes par classe		
1-photosynthesis (380)	0,71%	45	1,77%	1,1	2,03E-07
2-major CHO metabolism (188)	0,35%	11	0,43%	0,1	4,91E-01
3-minor CHO metabolism (215)	0,40%	16	0,63%	0,2	8,05E-02
4-glycolysis (117)	0,22%	4	0,16%	-0,1	6,64E-01
5-fermentation (58)	0,11%	2	0,08%	0,0	1,00E+00
6-gluconeogenese/ glyoxylate cycle (32)	0,06%	0	0,00%	-0,1	4,02E-01
7-oxidative pentose phosphate cycle (41)	0,08%	2	0,08%	0,0	7,21E-01
8-TCA / org. Transformation (125)	0,23%	5	0,20%	0,0	1,00E+00
9-mitochondrial electron transport / ATP synthesis (143)	0,27%	8	0,31%	0,0	5,57E-01
10-cell wall (597)	1,11%	45	1,77%	0,7	3,95E-03
11-lipid metabolism (574)	1,07%	42	1,65%	0,6	8,26E-03
12-N-metabolism (66)	0,12%	3	0,12%	0,0	1,00E+00
13-amino acid metabolism (405)	0,75%	28	1,10%	0,3	6,16E-02
14-S-assimilation (20)	0,04%	2	0,08%	0,0	2,62E-01
15-metal handling (83)	0,15%	10	0,39%	0,2	9,31E-03
16-secondary metabolism (679)	1,26%	35	1,37%	0,1	5,86E-01
17-hormone metabolism (743)	1,38%	58	2,28%	0,9	5,56E-04
18-Co-factor and vitamine metabolism (84)	0,16%	4	0,16%	0,0	1,00E+00
19-tetrapyrrole synthesis (70)	0,13%	21	0,82%	0,7	5,47E-10
20-stress (1418)	2,63%	55	2,16%	-0,5	1,61E-01
21-redox (269)	0,50%	30	1,18%	0,7	5,81E-05
22-polyamine metabolism (27)	0,05%	2	0,08%	0,0	3,79E-01
23-nucleotide metabolism (239)	0,44%	8	0,31%	-0,1	4,41E-01
24-biodegradation of Xenobiotics (37)	0,07%	5	0,20%	0,1	3,99E-02
25-C1-metabolism (56)	0,10%	2	0,08%	0,0	1,00E+00
26-miscellaneous enzyme families (2121)	3,94%	167	6,55%	2,6	1,43E-09
27-RNA (3053)	5,67%	163	6,40%	0,7	1,26E-01
28-DNA (750)	1,39%	22	0,86%	-0,5	2,29E-02
29-protein (4891)	9,08%	244	9,58%	0,5	3,98E-01
30-signalling (1952)	3,62%	112	4,40%	0,8	4,57E-02
31-cell (926)	1,72%	60	2,35%	0,6	2,01E-02
33-development (736)	1,37%	33	1,30%	-0,1	8,61E-01
34-transport (1468)	2,73%	100	3,92%	1,2	3,79E-04
35-not assigned (31287)	58,10%	1204	47,25%	-10,8	6,25E-27
TOTAL (53580)	100,00%	2548	100,00%		

Annexe II-7 - Répartition des sondes au sein des classes fonctionnelles dans l'analyse des gènes significativement différentiellement exprimés chez le mutant *dash* EMS109 par rapport au WT à 8 j. Les différences de représentativité des classes fonctionnelles ont été calculées afin de déterminer les classes ayant été enrichies (rouge) ou appauvries (bleu) en gènes. Les valeurs p associées à ces différences (test de Fisher exact) sont indiquées et un code couleur est attribué en fonction des valeurs.

	p < 0,001
	p < 0,01
	p < 0,05



Annexe II-8 – Vue d’ensemble de la répartition des gènes dans les différentes classes fonctionnelles BIN d’après le logiciel Mapman. Chaque carré représente l’intensité d’expression d’un gène de cette classe. Les dégradés de couleurs sont fonction des intensités des \log_2 ratio *dash* EMS109/WT. En rouge les valeurs des \log_2 ratio positifs indiquent les gènes sur-exprimés chez le mutant *dash* EMS109 et en bleu, les valeurs des \log_2 ratio négatifs représentent les gènes sous-exprimés. Les codes BIN sont décrits dans l’annexe II-7.

CHAPITRE III

-

*Etude du transcriptome des tissus
composant la graine au stade de
transition entre l'embryogenèse et
le remplissage*

III.1 - Introduction

Chez *Medicago truncatula* (*M. truncatula*), plante modèle pour l'étude des légumineuses, les graines sont confinées tout au long de leur développement dans des gousses, qui assurent à la fois un rôle de protection et de transfert de nutriments. La gousse dérive des tissus maternels, tout comme le tégument de la graine, alors que l'embryon et l'albumen sont issus de la double fécondation. Celle-ci fait intervenir deux gamètes mâles : l'un va féconder l'oosphère et le second va fusionner avec les noyaux polaires de la cellule centrale. La première fécondation évolue ensuite en un zygote diploïde qui deviendra l'embryon ; la deuxième conduira à la formation d'un zygote triploïde, à l'origine de l'albumen (Reiser and Fischer, 1993). Ces trois tissus évoluent ensuite conjointement tout au long du développement de la graine, caractérisée par trois étapes distinctes : l'embryogénèse, le remplissage et la maturation et enfin la dessiccation. L'embryogénèse constitue une phase riche en divisions et différenciations cellulaires, et notamment pour l'embryon qui vient d'être formé (Laux et al., 2004). S'en suivent ensuite des étapes de synthèse de composés de stockage (protéines, lipides, polysaccharides..) principalement dans l'embryon. L'acquisition de la tolérance à la dessiccation et de la vigueur germinative interviennent également pendant la phase de remplissage et de maturation (Gallardo et al., 2003). En dernier lieu lors de la dessiccation, la perte en eau que subit la graine permet d'optimiser sa conservation et de lui conférer sa dormance.

De nombreuses interactions ont été mises en évidence entre les tissus de la graine via les gènes qui y sont exprimés et ces échanges expliquent ce développement coordonné de l'albumen, de l'embryon et du tégument. Ainsi, l'importance de la cellularisation de l'albumen pour le développement de l'embryon a été souligné (Hehenberger et al., 2012), de même que la source de nutriments qu'il constitue pour l'embryon (Melkus et al., 2009), ou encore l'influence du développement de l'albumen sur celui du tégument et de la taille de la graine (Garcia et al., 2005; Haughn and Chaudhury, 2005). Plusieurs gènes exprimés dans l'albumen ont été identifiés comme ayant des conséquences sur le développement d'un ou plusieurs tissus de la graine. La mutation d'un facteur de transcription (FT) bHLH ZHOUP1 (ZOU) ou RETARDED GROWTH OF EMBRYO (RGE), induit des anomalies dans la structure de l'albumen qui affecte la morphologie de l'embryon chez *A. thaliana* (Kondou et al., 2008; Yang et al., 2008). En effet, l'albumen persistant observé dans les graines de ces mutants empêche le développement de l'embryon, conduisant à des graines asséchées de petite taille. De même, une subtilisine spécifique de l'albumen des graines chez *M. truncatula* SBT1.1 serait impliquée dans le contrôle la taille de la graine en régulant probablement le nombre de cellules des cotylédons (D'Erfurth et al., 2012).

Dans cette étude, nous proposons de mettre en avant la spécificité tissulaire de la graine en utilisant une approche transcriptomique. Nous avons choisi de cibler le stade 12 jap en vue de démontrer la

ensemble des sondes la puce			tri n°2 18231 gènes et contigs ayant une expression spécifique					
		tri n°1	tégument	embryon	albumen	embryon+ tégument	albumen+ tégument	albumen+ embryon
gènes et contigs	93445	47691	2931	3374	4110	3477	3107	1232
miRNA	231	70						
transposons	8447	3512						
total	102123	51273						

tri n°3 9130 gènes et contigs ayant une expression préférentielle					
tégument	embryon	albumen	embryon+ tégument	albumen+ tégument	albumen+ embryon
1173	1588	2225	1507	1568	1069

soit 27361 gènes et contigs ayant une expression spécifique ou préférentielle					
tégument	embryon	albumen	embryon+ tégument	albumen+ tégument	albumen+ embryon
4104	4962	6335	4984	4675	2301

et 20146 gènes et contigs exprimés dans les 3 tissus

tri n°1: sélection des sondes exprimées dans les tissus de la graine (sondes considérées comme non exprimées lorsque $\log_2 I < 1$ pour les 3 tissus)

tri n°2: sélection des sondes ayant une expression spécifique à un ou plusieurs tissus (soit au moins 2 fois plus exprimées dans un tissu donné, $\log_2 RI > |1|$ et $BH < 0,05$)

tri n°3: sélection des sondes préférentiellement exprimées dans un ou plusieurs tissus (soit au moins 1,5 fois plus exprimées dans un tissu donné, $\log_2 RI > |0,585|$ et $BH < 0,05$)

I: moyenne des intensités d'expression tissulaire pour les six répétitions (trois répétitions et deux comparaisons)

RI: ratio des moyennes des intensités d'expression tissulaire pour chaque comparaison

BH: valeur p de probabilité corrigée par un test Benjamini-Hochberg

Tableau III-1 –A. Nombres de sondes correspondants à des gènes ou contigs, miRNA et transposons présents sur la puce NimbleGen Medtr_v1.0. B. Répartition du nombre de gènes et contigs différentiellement exprimés dans les tissus de la graine.

contribution fonctionnelle de chaque tissu dans cette étape clé de transition entre l'embryogénèse (nombreuses divisions cellulaires au sein de l'embryon) et le remplissage (accumulation des réserves dans l'embryon) (Gallardo et al., 2007). Le transcriptome des trois tissus de la graine a donc été étudié et comparé à ce stade (étude réalisée en collaboration avec Julia Buitink, IRHS Angers).

III.2 - Résultats et discussion

III.2.1 - Stratégie d'étude du transcriptome des trois tissus de la graine

Pour cette étude, la puce NimbleGen Medtr_v1.0 (développée par l'équipe de Julia Buitink, UMR IRHS – Angers) a été utilisée afin de définir la signature d'expression génique des trois tissus à 12 jap. Au total 102 123 sondes de gènes dont on dispose de la séquence complète ou partielle (contigs), de transposons et de miRNAs sont représentées sur cette puce. Le nombre de sondes dans chacune de ces catégories est présenté dans le tableau III-1A. Les trois tissus ont été hybridés de manière à comparer chacun avec les deux autres : embryon vs albumen, tégument vs albumen et tégument vs embryon; chaque comparaison ayant été répétée trois fois (*i.e.* trois répétitions biologiques). Le fichier obtenu après traitement des données comporte pour chaque sonde, les intensités (I) dans chaque tissu exprimées en \log_2 ($\log_2 I$, moyenne des six répétitions), les \log_2 des ratios d'intensité ($\log_2 RI$, moyenne des trois répétitions) pour chaque comparaison embryon vs albumen, tégument vs albumen et tégument vs embryon et les valeurs p de probabilité corrigées par le test Benjamini-Hochberg. Nous avons ensuite analysé de manière approfondie les données de la catégorie de sondes "gènes et contigs" pour laquelle nous disposons d'une annotation fonctionnelle BIN (*i.e.* classes fonctionnelles) selon MapMan (Thimm et al., 2004).

Un premier tri a permis d'éliminer les gènes et contigs qui ne présentaient pas d'expression dans les tissus ($\log_2 I < 1$) et un second tri a révélé les gènes spécifiquement exprimés dans un ou plusieurs tissus [tableau III-1B]. Ce second tri a consisté à étudier les moyennes des intensités de chacune de ces sondes dans les trois tissus; en considérant qu'une sonde est spécifique d'un ou de deux tissus si $\log_2 RI > |1|$ (*i.e.* gènes au moins deux fois plus exprimés dans un tissu donné). Ainsi, 18 231 gènes ayant des spécificités tissulaires ont été révélés. Ces gènes sont répartis pour les analyses relatives à chacun des trois tissus pris individuellement, puis pour les tissus pris deux à deux [tableau III-1B]. Un troisième tri sur les gènes restants a permis d'identifier 9 130 gènes préférentiellement exprimés dans un ou plusieurs tissus ($\log_2 RI > |0,585|$, *i.e.* gènes au moins 1,5 fois plus exprimés dans un tissu donné). Suite à ce dernier tri, 20 146 gènes et contigs dont les valeurs ne permettent pas un tri en suivant les seuils précédemment énoncés, sont considérés comme exprimés dans les trois tissus.

Afin de tenir compte de l'ensemble des gènes exprimés dans la graine, puis de se focaliser sur ceux considérés comme spécifiques ou préférentiels, deux approches ont été utilisées : une approche globale

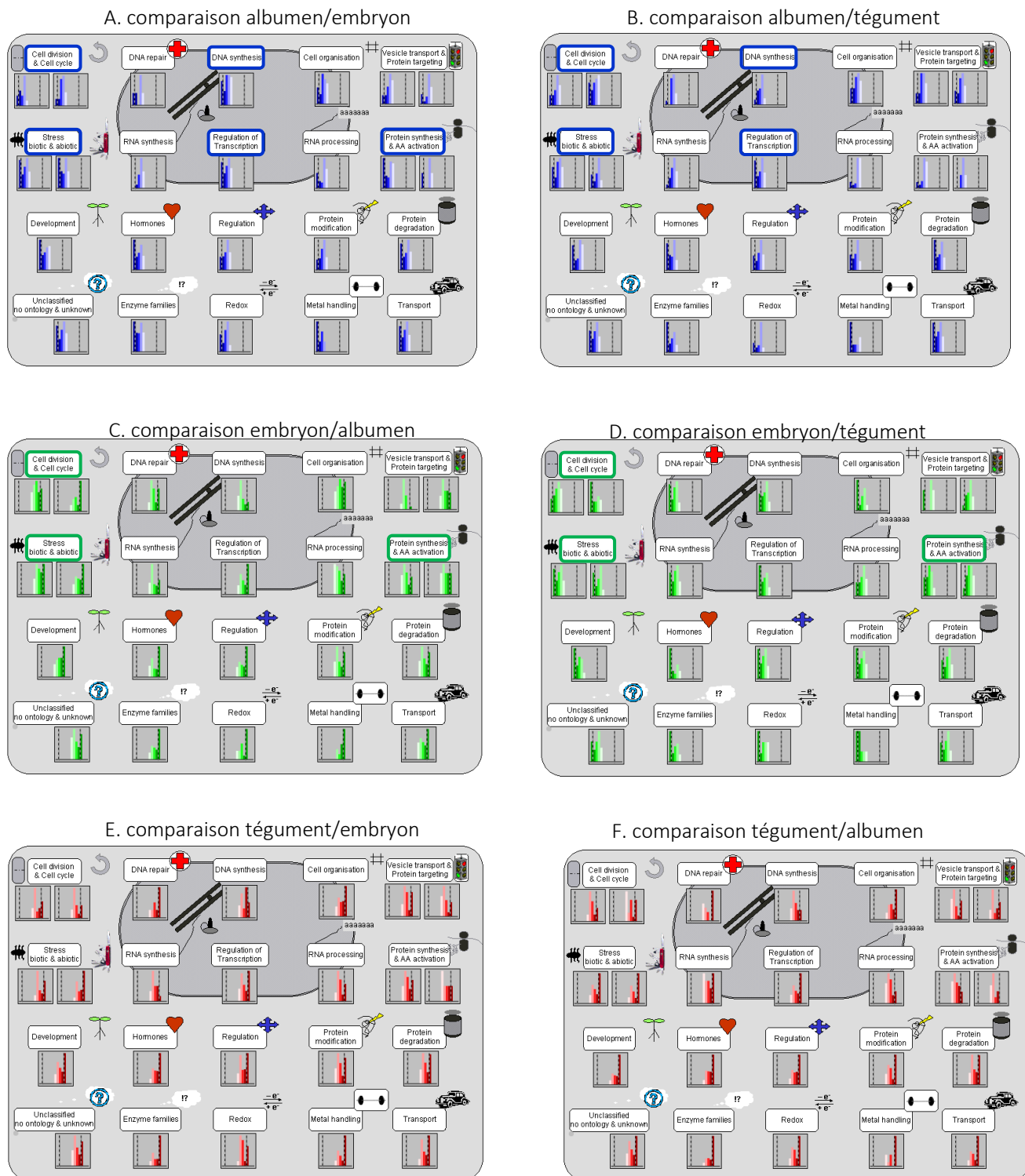


Figure III-1 – Analyse comparative des fonctions cellulaires d'après le logiciel MapMan. La hauteur des histogrammes est fonction du nombre de gènes et les intensités des couleurs sont relatives à l'expression plus ou moins fortes des gènes dans les tissus : vert : embryon ; bleu : albumen ; rouge : tégument).

(A) Gènes préférentiellement exprimés dans l'albumen d'après les comparaisons albumen/embryon et (B) albumen/tégument. (C) Gènes préférentiellement exprimés dans l'embryon d'après les comparaisons embryon/albumen et (D) embryon/tégument. (E) Gènes préférentiellement exprimés dans le tégument d'après les comparaisons tégument/embryon et (F) tégument/albumen.

prenant en compte l'ensemble des gènes exprimés dans la graine, soit 47 691 gènes (tri n°1), et une approche ciblée sur les 27 361 gènes considérés comme spécifiques ou préférentiellement exprimés dans un ou plusieurs tissus (tri n°2 et 3).

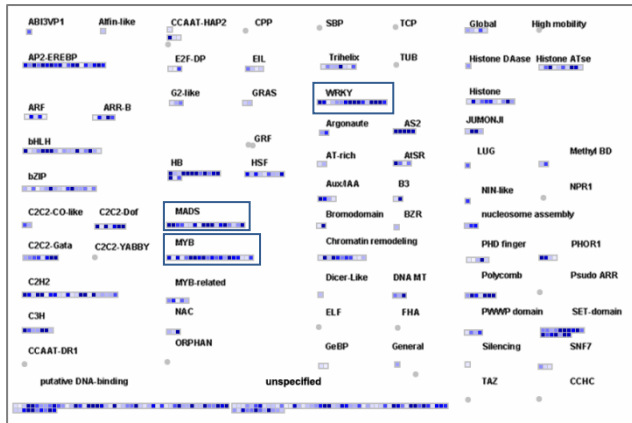
III.2.2 - Analyse globale portant sur l'ensemble des gènes exprimés dans les tissus de la graine

III.2.2.1 - Analyse globale des fonctions cellulaires

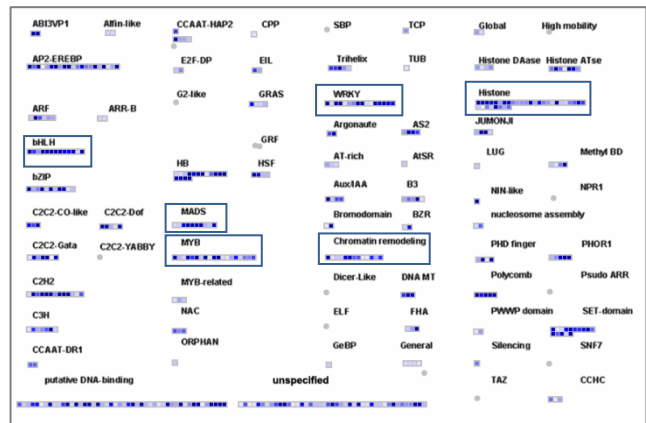
Cette analyse se base sur les $\log_2\text{RI}$ des trois comparaisons (embryon vs albumen, tégument vs albumen et tégument vs embryon) pour les 47 691 gènes et contigs. Elle est réalisée grâce au logiciel MapMan (Thimm et al., 2004) qui répertorie les gènes en catégories fonctionnelles et permet une visualisation globale des fonctions cellulaires. Cette analyse MapMan a été réalisée à partir des $\log_2\text{RI}$ pour lesquels la valeur p associée est inférieure à 0,05 ; soit 28 120 gènes pour la comparaison embryon vs albumen, 23 251 pour la comparaison tégument vs albumen et 21 812 pour la comparaison tégument vs embryon. L'annexe III-1 présente une vue d'ensemble des données classées selon MapMan, pour les trois comparaisons (albumen vs embryon, albumen vs tégument et embryon vs tégument) avec un code couleur relatif à chaque tissu (vert : gènes plus exprimés dans l'embryon ; bleu : gènes plus exprimés dans l'albumen ; rouge : gènes plus exprimés dans le tégument). Il est à noter que seuls les gènes pour lesquels un code MapMan a pu être attribué sont représentés sur les cartes (le nombre de gènes est indiqué pour chaque comparaison). La figure III-1 présente une vue d'ensemble des fonctions cellulaires pour chaque tissu avec le même code couleur. Deux fichiers images sont générés pour chaque tissu étant donné que chaque tissu est comparé aux deux autres. Il n'est pas opportun de fusionner les données afin de ne générer qu'une seule image par tissu car c'est au risque de perdre les valeurs des $\log_2\text{RI}$ de l'une des deux combinaisons lorsque les mêmes gènes sont différenciellement exprimés entre deux tissus pour les deux combinaisons.

Globalement, les résultats montrent que les activités cellulaires liées à l'activation des acides aminés et à la synthèse protéique se déroulent préférentiellement dans l'embryon [figures III-1C et D, annexe III-1A et C], qui est en forte multiplication cellulaire durant l'embryogénèse. On retrouve d'ailleurs que les divisions cellulaires sont très marquées dans ce tissu [figures III-1C et D et annexe III-1C]. Dans l'albumen, les gènes impliqués dans la régulation de la transcription, la synthèse d'ADN et le contrôle du cycle cellulaire sont fortement représentés [figures III-1A et B et annexe III-1B]. De façon surprenante, les gènes liés aux stress biotiques et abiotiques semblent fortement exprimés dans l'albumen et l'embryon [figures III-1A, B, C et D] par rapport au tégument. De façon générale, les gènes exprimés dans le tégument montrent une répartition homogène pour la majorité des fonctions cellulaires présentées sur ces cartes ; avec un nombre important de gènes fortement exprimés mais peu de gènes avec des niveaux d'expression

A. comparaison albumen/embryon



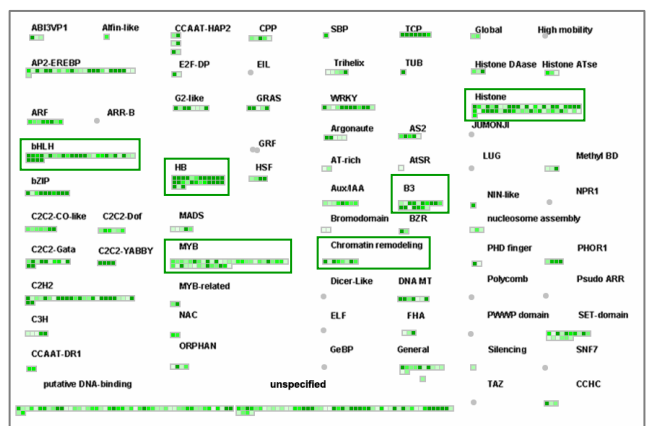
B. comparaison albumen/tégument



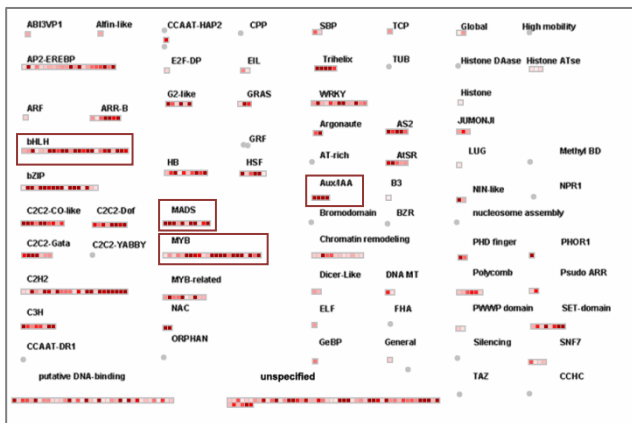
C. comparaison embryon/albumen



D. comparaison embryon/tégument



E. comparaison tégument/embryon



F. comparaison tégument/albumen

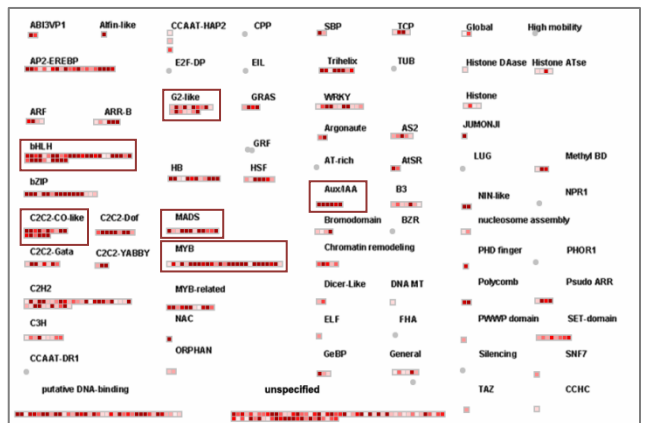


Figure III-2 – Analyse comparative des cartes "transcription" générées par le logiciel MapMan. Chaque carré de couleur représente un gène et les intensités des couleurs sont relatives à l'expression plus ou moins fortes des gènes dans les tissus : vert : embryon ; bleu : albumen ; rouge : tégument.

(A) Gènes préférentiellement exprimés dans l'albumen d'après les comparaisons albumen/embryon et (B) albumen/tégument. (C) Gènes préférentiellement exprimés dans l'embryon d'après les comparaisons embryon/albumen et (D) embryon/tégument. (E) Gènes préférentiellement exprimés dans le tégument d'après les comparaisons tégument/embryon et (F) tégument/albumen.

intermédiaires. Il est donc difficile de déterminer des spécificités dans les fonctions cellulaires attribuées au tégument.

III.2.2.2 - Analyse des cartes "transcription" générées par le logiciel MapMan

Les cartes "transcription" obtenues grâce au logiciel MapMan représentent le nombre de gènes impliqués dans la régulation de la transcription pour les trois comparaisons de tissus [annexe III-2]. Le nombre de gènes représentés sur ces cartes est indiqué pour chaque comparaison. Le code couleur est identique à celui décrit précédemment : bleu pour l'albumen, rouge pour le tégument et vert pour l'embryon. La figure III-2 présente cette même vue d'ensemble de la répartition des FTs pour chaque tissu comparé aux deux autres.

Les gènes relatifs aux histones et au remodelage de la chromatine sont préférentiellement exprimés dans l'albumen et l'embryon, comparé au tégument, suggérant une forte activité transcriptionnelle dans ces tissus [figures III-2B et D]. La classe des FTs de type B3 est majoritairement représentée dans l'embryon [figures III-2C et D], ce qui a également été observé dans la graine de maïs à 9 jap, stade correspondant à la transition embryogénèse-remplissage (Lu et al., 2013). Des FTs de la famille B3 tel que LEC2 et FUS3 ont été décrits comme associés au développement de l'embryon (Bäumlein et al., 1994; Stone et al., 2001; Agarwal et al., 2011). Chez *M. truncatula*, des FTs de type FUS3-like et ABI3-like ont un profil d'expression préférentiel de la fin d'embryogénèse, suggérant ainsi un rôle dans le développement de l'embryon (Kurdyukov et al., 2014). On retrouve également une importante proportion de FTs de type HB dans l'embryon [figures III-2C et D], ces FTs étant impliqués dans le développement et la différenciation cellulaire avec une expression très précoce lors des premières divisions cellulaires de l'embryon ou en fin d'embryogénèse (Nardmann et al., 2007; Wu et al., 2007; Kurdyukov et al., 2014). Ces familles de FT (*i.e.* B3 et HB) seraient donc fortement liées au développement de l'embryon.

Dans l'albumen, nous observons une expression préférentielle des FTs de type WRKY [figures III-2A et B], décrits comme impliqués dans la régulation de la taille de la graine (Garcia et al., 2003; Luo et al., 2005), et dans le tégument, une expression préférentielle des FTs relatifs aux auxines Aux/IAA [figures III-2E et F].

On constate que plusieurs familles de FTs, comme les G2-like ou les C2C2-CO-like ont des profils d'expression préférentiels de l'embryon et du tégument [figures III-2C et F]. Ceux de type bHLH sont aussi fortement exprimés dans le tégument et l'embryon [figures III-2C et F], bien que l'on en retrouve une dizaine fortement exprimés dans l'albumen [figure III-2B]. On note également que les facteurs de type MADS sont préférentiellement exprimés dans l'albumen et le tégument [figures III-2A, B, E et F], comme précédemment observé lors d'une étude de la distribution d'ESTs représentatifs de FTs dans les tissus de la graine de lin au cours de son développement (Venglat et al., 2011). Cette même étude a montré une prédominance des FTs de type MYB dans le tégument au stade cotylédons linéaires. Les cartes MapMan montrent de façon similaire que les gènes codant les facteurs de type MYB sont plus fortement exprimés dans le tégument [figures III-2E et F].

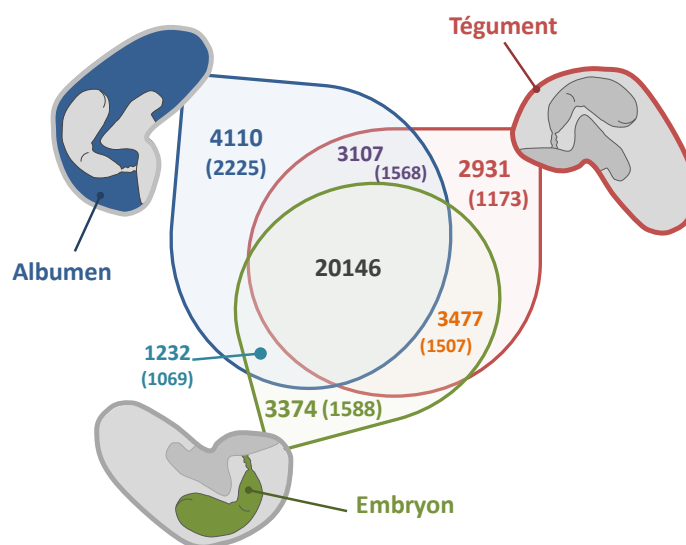


Figure III-3 – Diagramme représentatif du nombre de gènes spécifiquement et préférentiellement (indiqué entre parenthèses) exprimés dans chacun des trois tissus de la graine, suite aux tris n°2 et 3 [tableau III-1].

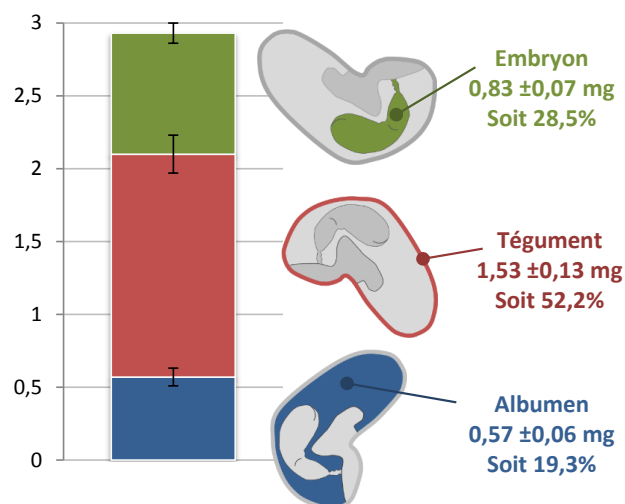


Figure III-4 – Composition massique (en mg) des tissus d'une graine à 12 jap.

III.2.3 - Analyse ciblée des gènes préférentiellement exprimés dans un tissu

III.2.3.1 - Analyse des spécificités tissulaires

Dans cette 2^{ème} approche ciblée, nous avons étudié plus spécifiquement la répartition tissulaire de l'ensemble des gènes exprimés dans la graine grâce aux tris n°2 et 3 décrits précédemment (cf. paragraphe III.2.1 - Stratégie d'étude du transcriptome des trois tissus de la graine" et tableau III-1). Cette répartition est présentée dans la figure III-3 et tient compte du nombre de gènes spécifiquement ou préférentiellement exprimés pour chaque tissu. Conformément aux nombres de sondes présentes sur la puce, l'albumen constitue le tissu pour lequel on retrouve le plus de gènes qui lui sont spécifiques (*i.e.* 4 110), alors que l'embryon et le tégument ont une spécificité tissulaire en termes de nombre de gènes moindre, avec respectivement 3 374 et 2 931 gènes. Cette répartition reste identique que l'on tienne compte des gènes spécifiques et/ou préférentiels, l'albumen restant le tissu le plus riche en gènes devant l'embryon et le tégument. De façon surprenante, une part importante de gènes se retrouvent exprimés dans ces deux tissus (*i.e.* embryon et tégument) 3 477 gènes spécifiques et 1 507 préférentiels, alors qu'une étude du transcriptome de la graine à différents stades de développement chez *A. thaliana* a montré que le tégument diffère de l'embryon et de l'albumen qui partagent quant à eux de nombreuses similitudes (Belmonte et al., 2013). Cette différence pourrait provenir de contamination lors de la préparation des échantillons. En effet pour *M. truncatula*, les tissus ont été isolés manuellement sous loupe binoculaire et il est possible que les couches d'albumen périphériques soient restées accolées au tégument ou à l'embryon, les contaminations embryon-tégument sont quant à elles moins probables. En effet, les transcrits du facteur de transcription de type DOF (Medtr2g014060.1) spécifiques des couches de cellules les plus accolées au tégument (cf article en préparation, chapitre 2), sont retrouvés dans l'albumen (\log_2 albumen = 4,41) et le tégument (\log_2 tégument = 2,75). Les analyses chez *A. thaliana* proviennent d'une technique de dissection plus précise avec la microdissection laser et les tissus analysés sont également beaucoup plus nombreux et détaillés (albumen chalazal, périphérique..). De plus, l'analyse présentée ici chez *M. truncatula* est focalisée sur un seul stade de développement (*i.e.* 12 jap) ; ce stade n'a pas été étudié dans Belmonte et al. (2013), où les analyses ciblent des stades plus jeunes (jusqu'au stade linéaire) ou plus tardif (graine mature) (Belmonte et al., 2013). L'étude comparative du transcriptome de l'embryon vs albumen chez le maïs à 9 jap, stade correspondant à la transition embryogenèse-remplissage, a montré comme chez *A. thaliana* une part importante de gènes exprimés à la fois dans l'embryon et l'albumen, malgré un plus grand nombre de transcrits détectés dans l'embryon (Lu et al., 2013). Il est à noter que le tégument n'a pas été étudié dans cette analyse transcriptomique. Les fonctions attribuées aux tissus de la graine chez ces deux espèces sont différentes, et particulièrement l'évolution de l'albumen, transitoire chez *M. truncatula* et persistant chez le maïs. De même, la morphologie et chronologie du développement de la graine sont distinctes. On peut donc en conclure que les spécificités tissulaires de la graine varient en fonction de la structure du tissu au cours des stades de développement et diffèrent entre espèces.

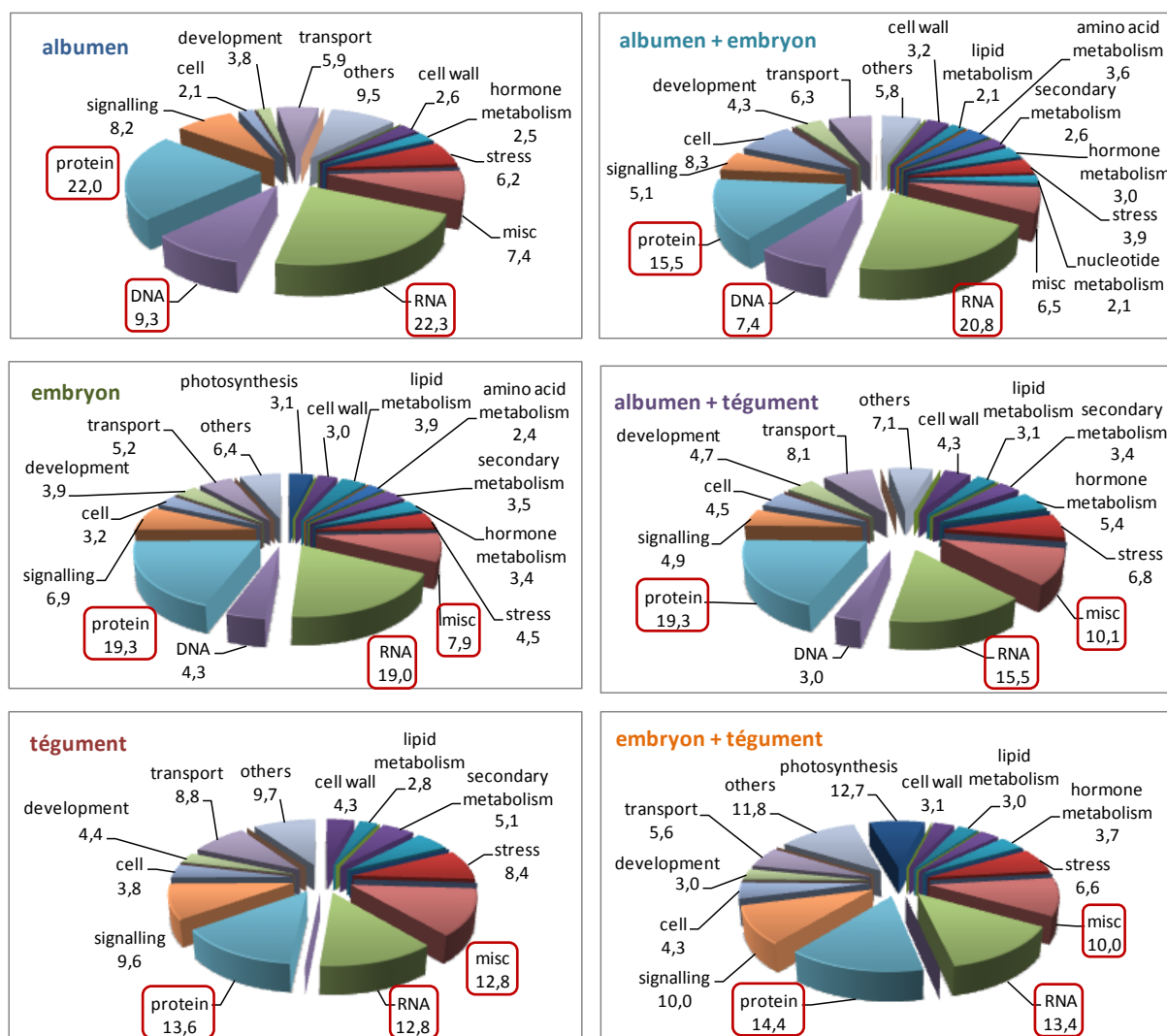


Figure III-5 – Classification fonctionnelles (BIN code) des gènes spécifiques de chaque tissu : tégument, embryon et albumen et des gènes spécifiquement exprimés dans deux tissus : embryon et tégument, albumen et embryon et albumen et tégument. Ces diagrammes ne font pas apparaître la classe fonctionnelle "35.not assigned", qui n'a pas non plus été prise en compte pour le calcul des pourcentages de représentativité. Les classes dont les pourcentages sont inférieurs à 2% ont été regroupées dans la classe "others".

Il est intéressant de comparer les données transcriptomiques avec la part relative de chaque tissu dans la graine, estimée grâce aux pesées réalisées lors de la préparation des échantillons [annexe III-3]. Les trois tissus ont été séparés à partir de plusieurs échantillons de 50 à 100 graines. La masse de chaque tissu ramenée à une graine a ensuite été calculée et la composition massique des tissus d'une graine à 12 jap a pu être estimée [figure III-4]. La conversion de ces masses en pourcentage nous a permis d'estimer qu'à 12 jap, le tégument représente 52% du poids d'une graine ; l'embryon 29% et l'albumen seulement 19%, (valeurs comparables à celles présentées dans l'étude de plusieurs génotypes à 13 jap de Djemel et al. 2005 (Djemel et al., 2005)). Ces valeurs sont inversement corrélées à la part que représente chaque tissu dans la structure de la graine à 12 jap, où l'albumen occupe le plus "d'espace", suivi du tégument et enfin de l'embryon. L'albumen reste cependant le tissu le plus riche en termes de diversité de gènes transcrits, dérivant probablement de son rôle important dans le stockage temporaire des réserves et le transfert de nutriments. Ceci peut également s'expliquer par le fait que c'est un tissu qui évolue rapidement tout au long du développement de la graine, d'où cette forte diversité d'activité transcriptionnelle révélée. En effet, l'albumen est tout d'abord un tissu syncytial qui se cellularise puis disparaît progressivement au profit du développement de l'embryon.

III.2.3.2 - Analyse des classes fonctionnelles

Suite à la répartition des transcrits au sein de chaque tissu [figure III-3], il est intéressant de se tourner vers l'identification et l'étude des classes fonctionnelles. Les correspondances avec les annotations fonctionnelles BIN dérivées de MapMan (Thimm et al., 2004) ont été recherchées pour l'ensemble des gènes exprimés spécifiquement et préférentiellement dans chaque tissu ou dans plusieurs tissus. La répartition de ces classes fonctionnelles, exprimée en pourcentage, est représentée respectivement pour les gènes spécifiquement exprimés, puis pour les gènes spécifiques et préférentiels, sous forme de diagramme dans les annexes III-4 et III-5. Les classes fonctionnelles pour lesquelles la représentation est inférieure à 2% ont été regroupées dans la classe "others". La classe "35.not assigned" est la plus représentée quelque soit les analyses, de 33% pour embryon+tégument [annexe III-4] à 62% pour albumen+embryon+tégument [annexe III-5].

Dans un souci de clarté d'analyse, nous avons recalculé ces pourcentages de répartition en retirant cette classe "35.not assigned" afin de mieux visualiser les classes les plus fortement représentées [figures III-5 et III-6]. De façon générale, quelque soit l'analyse présentée (ie "gènes spécifiques" ou "gènes spécifiques et préférentiels") et le tissu, les classes les plus représentées sont "27.RNA", "29.protein" et "26.misc". L'analyse de ces trois classes sera détaillée par la suite. La classe "28.DNA" est également retrouvée parmi les plus représentées dans l'albumen et albumen+embryon mais aussi dans l'analyse commune aux trois tissus. L'analyse détaillée des gènes de cette classe indique une forte proportion de gènes liés à la synthèse des histones et au remodelage de la chromatine dans l'albumen et l'embryon, comme

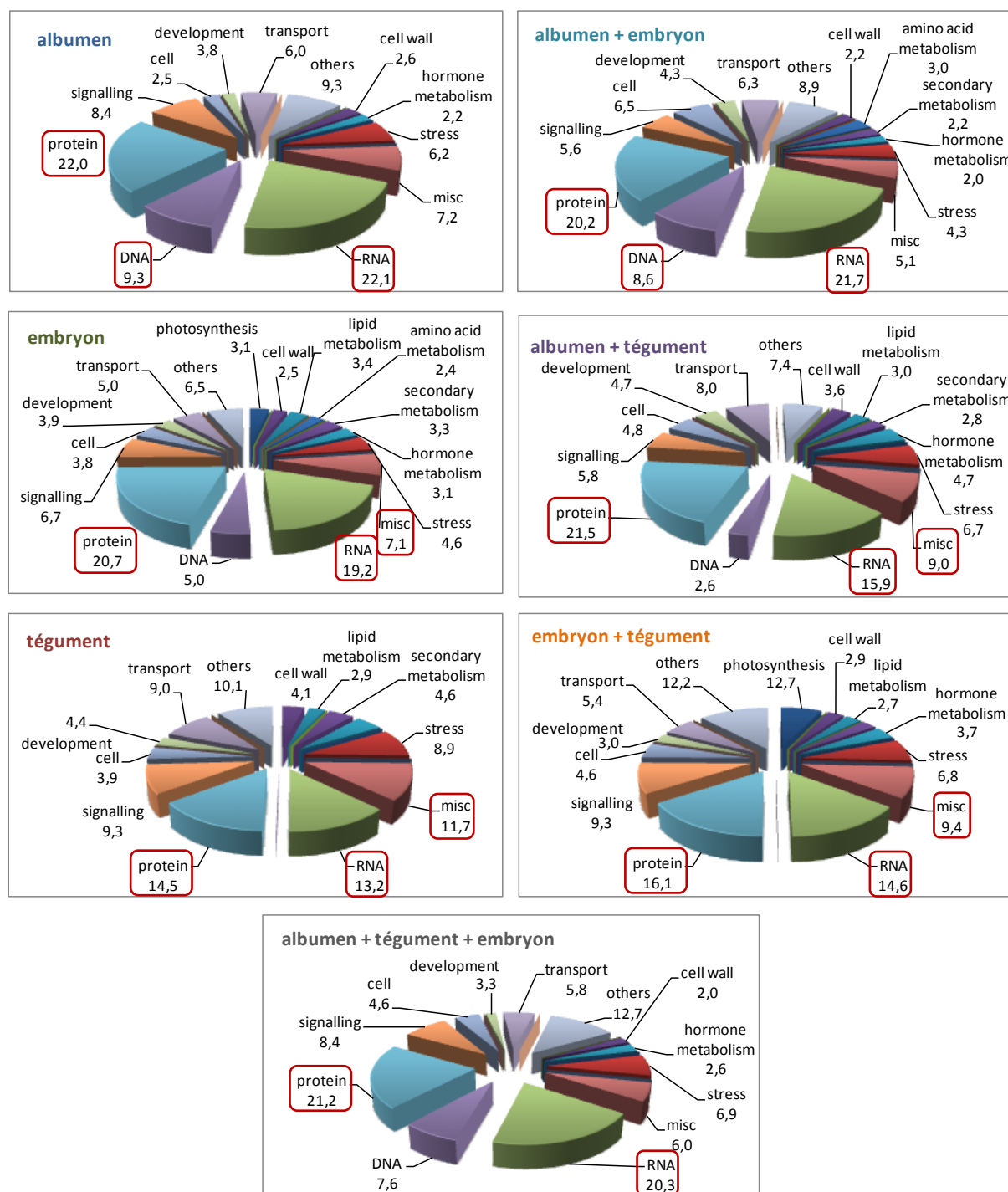


Figure III-6 – Classification fonctionnelle (BIN code) des gènes spécifiques et préférentiels de chaque tissu : tégument, embryon et albumen et des gènes spécifiquement ou préférentiellement exprimés dans deux tissus : embryon et tégument, albumen et embryon et albumen et tégument ou dans les trois tissus. Ces diagrammes ne font pas apparaître la classe fonctionnelle "35.not assigned", qui n'a pas non plus été prise en compte pour le calcul des pourcentages de représentativité. Les classes dont les pourcentages sont inférieurs à 2% ont été regroupées dans la classe "others".

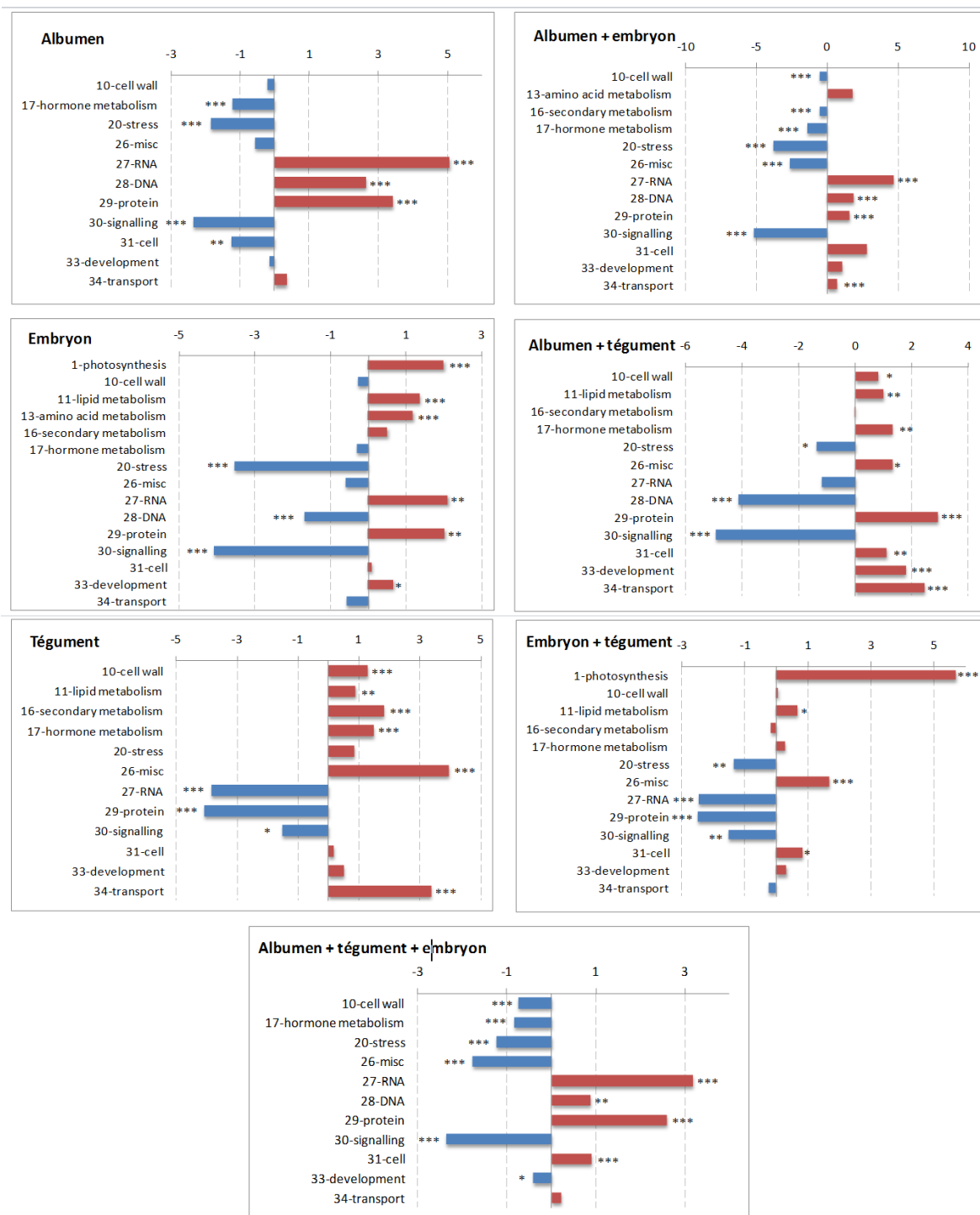


Figure III-7 – Histogramme de sur- (rouge) ou sous- (bleu) représentation des classes fonctionnelles BIN révélées par l'analyse transcriptomique des gènes spécifiques et préférentiels de chaque tissu : tégument, embryon et albumen et des gènes communs à deux ou aux trois tissus. Ces représentations ne font pas apparaître la classe fonctionnelle "35.not assigned", qui n'a pas non plus été prise en compte pour les calculs. Les classes dont les pourcentages sont inférieurs à 2% ne sont pas représentées. Les valeurs sont obtenues par la différence des pourcentages de représentativité des classes BIN entre les analyses par tissu et la totalité des sondes présentes sur la puce [tableau annexe III-7]. Les valeurs p de probabilité sont indiquées : *** p<0,001 ; **p<0,01, *p<0,05.

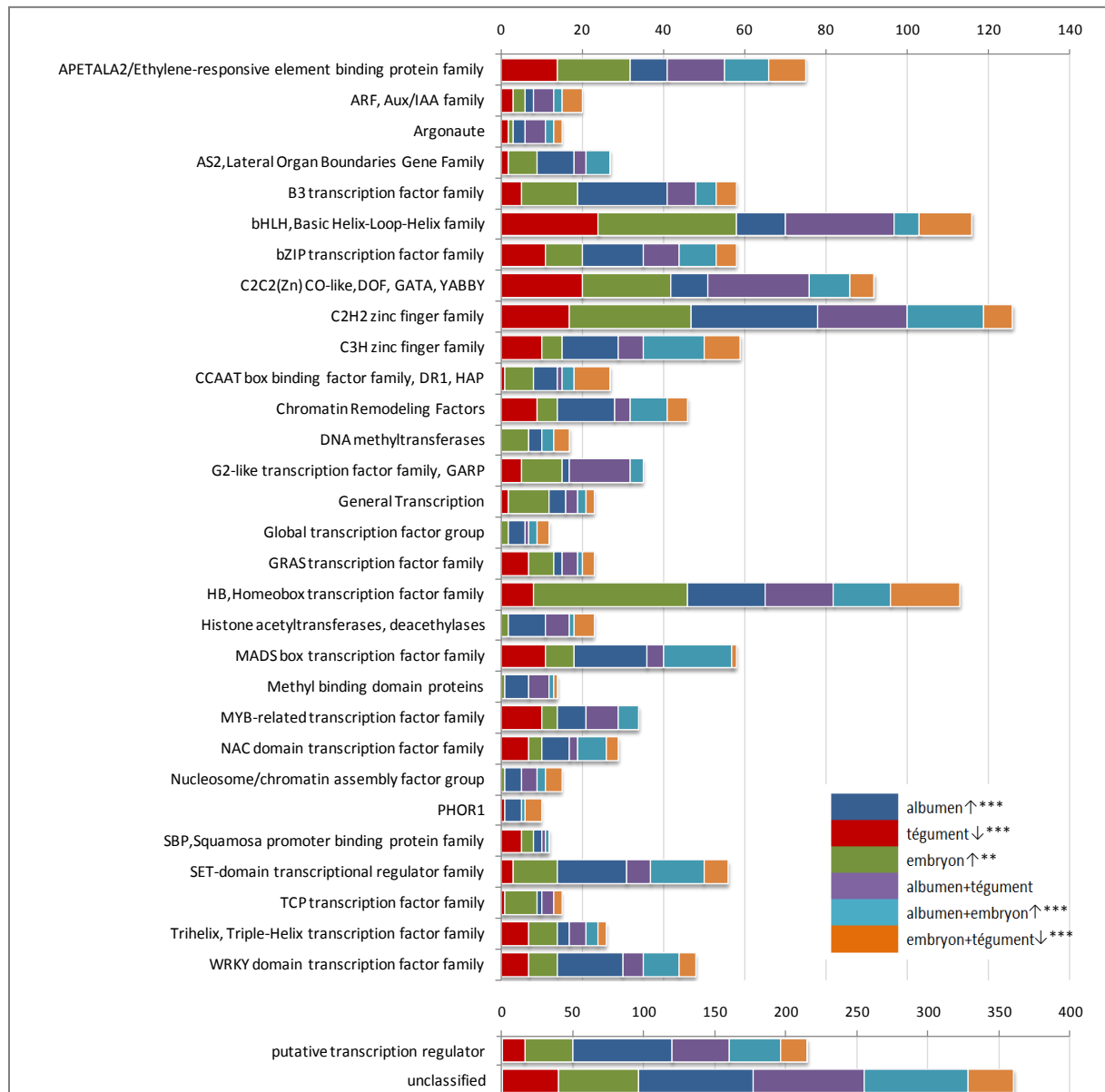


Figure III-8 – Nombre de gènes (ou contigs) dans les différentes familles de la classe BIN "27.3.RNA.regulation of transcription". La distribution de ces gènes dans les trois tissus (expression spécifique ou préférentielle) est présentée par un code couleur. Les familles comportant moins de 10 gènes ne sont pas représentées sur cette figure. Les symboles ↑ et ↓ correspondent respectivement aux familles sur- et sous-représentées dans le(s) tissu(s) d'après la figure III-7 et l'annexe III-7 avec les valeurs p de probabilité : *** p<0,001 ; **p<0,01, *p<0,05.

précédemment observé lors de l'étude des gènes liés à la transcription (paragraphe III.2.2.2. - Analyse des cartes "transcription" générées par le logiciel MapMan et figures III-2B et D).

Les informations relatives à la distribution des classes fonctionnelles sont cependant à modérer en fonction de la répartition des gènes au sein de ces classes pour la totalité des sondes présentes sur la puce. En effet cette répartition, présentée dans l'annexe III-6A et B, indique également les classes "29.protein" et "27.RNA" comme les plus représentées. Il convient alors de tenir compte de cette répartition pour estimer plus justement les régulations biologiques au sein de chaque tissu. Dans ce but, une valeur de sur- ou sous-représentation, obtenue par la différence des pourcentages de répartition, a donc été calculée pour chaque classe fonctionnelle BIN de l'analyse pour un, deux ou trois tissus, et ce par rapport à la représentation de la classe pour l'ensemble de la puce, en ne tenant pas compte de la classe "35.not assigned" [annexe III-7]. Les histogrammes relatifs à ces valeurs de sur ou sous-représentation sont présentés pour chaque analyse tissulaire dans la figure III-7 pour les classes fonctionnelles dont la représentation est supérieure à 2%; le détail des valeurs pour l'ensemble des classes est présenté dans l'annexe III-7. En tenant compte de ce paramètre de sur ou sous-représentation, la classe fonctionnelle "27.RNA", qui était la plus représentée pour tous les tissus reste donc la plus présente uniquement pour les analyses des tissus albumen et/ou embryon. Ces données renforcent l'hypothèse soulevée précédemment (suite à l'analyse globale des facteurs de régulation) d'une activité transcriptionnelle intense dans ces tissus. On note également une forte représentativité de cette classe "27.RNA" dans l'analyse des gènes communs aux trois tissus. Alors que la classe "29.protein" semblait fortement représentée pour l'ensemble des tissus [figure III-6], elle ne montre une sur-représentation significative que pour les analyses des tissus albumen et/ou embryon, et pour l'analyse des gènes communs aux trois tissus [figure III-7]. Quant à la classe "26.misc", l'analyse des sur- et sous- représentations conforte la part importante qu'occupe cette classe fonctionnelle pour les analyses : tégument, tégument et embryon, albumen et tégument.

a - Classe fonctionnelle "27.RNA"

Comme indiqué précédemment, la classe la plus représentée dans l'albumen et/ou l'embryon et albumen+tégument+embryon est la classe "27.RNA". L'analyse détaillée de cette classe selon le découpage BIN indique une forte proportion de FTs. Les gènes codant les FTs sont classés dans la classe BIN "27.3.RNA.regulation of transcription" où l'on retrouve la majorité des gènes de la classe "27.RNA" (par rapport aux autres BIN codes: "27.3.1.RNA.processing", "27.3.2.RNA.transcription" ou encore "27.4.RNA.binding"). Le détail des différentes familles de FTs et leur distribution dans les trois tissus (expression spécifique et préférentielle) sont présentés dans la figure III-8. Les familles qui ne sont représentées que par un seul gène n'apparaissent pas dans cette figure. Bien que cette classe ne soit pas sur-représentée dans le tégument, les données relatives à ce tissu ont été également prises en compte afin d'observer la répartition d'expression des gènes dans les trois tissus.

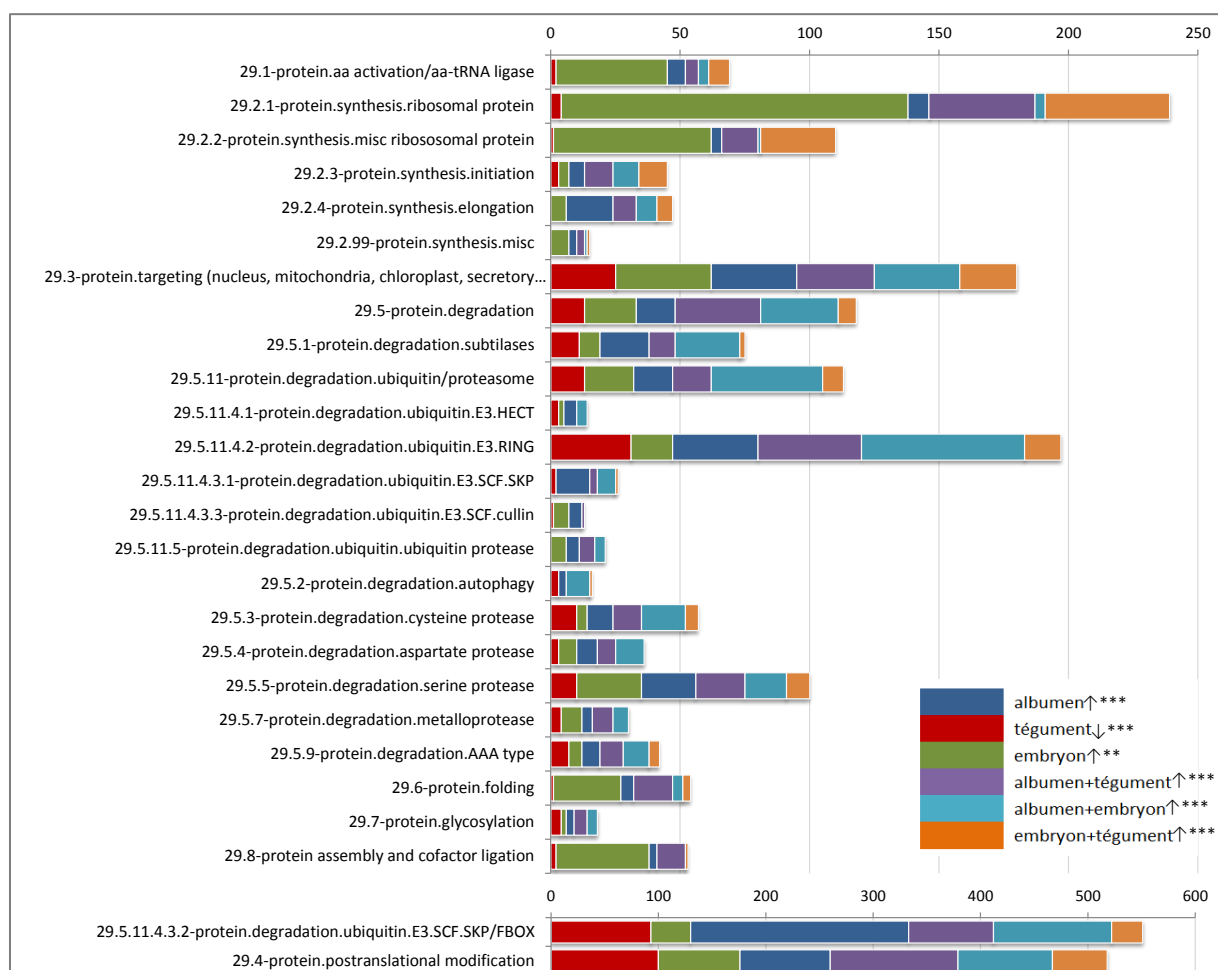


Figure III-9 – Nombre de gènes (ou contigs) dans les différentes familles de la classe BIN "29.protein". La distribution de ces gènes dans les trois tissus (expression spécifique ou préférentielle) est présentée par un code couleur. Les familles comportant moins de 10 gènes ne sont représentées sur cette figure. Les symboles ↑ et ↓ correspondent respectivement aux familles sur- et sous-représentées dans le(s) tissu(s) d'après la figure III-7 et l'annexe III-7 avec des valeurs p de probabilité : *** p<0,001 ; **p<0,01, *p<0,05.

Les FTs les plus représentés sont les zinc finger (C2H2 et C2C2), les FTs de type bHLH et HB que l'on retrouve de façon prédominante dans l'embryon. Deux autres classes de FTs aux fonctions indéterminées ("putative transcription regulator" et "unclassified") sont très fortement représentées, majoritairement au niveau de l'albumen. L'analyse des gènes de ces classes *via* la recherche des identifiants dans la nouvelle version du génome Mt4.0 n'apporte pas plus d'informations en termes de description. Globalement, les résultats de cette approche ciblée sont conformes à ceux révélés par l'analyse globale des FTs (cf. paragraphe III.2.2.2-Analyse des cartes "transcription" générées par le logiciel MapMan et figure III-2) qui suggérerait par exemple la prédominance des FTs de type HB dans l'embryon, des FTs WRKY dans l'albumen ou encore la spécificité d'expression des FTs de types MADS dans l'albumen et le tégument.

b - Classe fonctionnelle "29.protein"

La classe "29.protein" est sur-représentée dans l'albumen et l'embryon [figure III-7]. Les protéines impliqués dans la synthèse des ribosomes sont fortement représentées dans l'embryon (expression spécifique ou préférentielle), cette catégorie ressortait déjà de façon prédominante dans l'embryon d'après les cartes "fonctions cellulaires" de l'analyse globale *via* MapMan (cf paragraphe III.2.2.1 et figure III.1). Parmi les gènes de cette classe "29.protein" spécifiquement ou préférentiellement exprimés dans l'albumen, on retrouve une part importante de gènes liés à la dégradation des protéines ("29.5.protein.degradation" 366/540) et plus précisément à la dégradation *via* l'ubiquitine ("29.5.11.protein.degradation.ubiquitin" 279/366) [figure III-9]. Les séquences des gènes et contigs de cette classe 29.5.11 sont en grande partie annotées comme des protéines F-box/kelch-repeat : le domaine F-box en N-terminal permet une interaction avec les protéines Skp1 et la partie C-terminal comportant les motifs kelch (séquences répétées) reconnaît les protéines cibles. La fonction principale de ces protéines est de s'associer à d'autres protéines en vue de les marquer afin pour qu'elles soient détruites par des voies de dégradation dépendantes de l'ubiquitine. Plusieurs études ont montré que cette régulation est impliquée dans de nombreux processus de développement chez les plantes, comme la sénescence des feuilles et le contrôle de la taille des grains chez le riz (Chen et al., 2013). Chez le riz, une protéine contenant le domaine F-box et les motifs kelch est codée par un gène *OsFBK12* préférentiellement exprimé dans les panicules et grains. Elle a été décrite comme intervenant dans la régulation négative de la synthèse d'éthylène *via* son interaction avec les protéines SAMS (S-Adenosyl-L-Methionine Synthetase) qui se traduit par un retard de la sénescence des feuilles. Des mutants RNAi et sur-expresser ont été générés pour ce gène *OsFBK12*. Le poids de mille grains diminue significativement chez les mutants où l'expression de ce gène est éteinte et augmente lorsqu'il est surexprimé. Cette étude suggère également qu'*OsFBK12* est impliquée dans l'architecture des panicules, et dans la morphologie des enveloppes et des granules d'amidon. Ces informations laissent penser que la protéine *OsFBK12* pourrait affecter les divisions cellulaires et influencer la taille de la graine (Chen et al., 2013). Parallèlement, cette étude a permis de mettre en évidence que les protéines de type F-box ont des profils d'expression qui varient au cours du développement des panicules

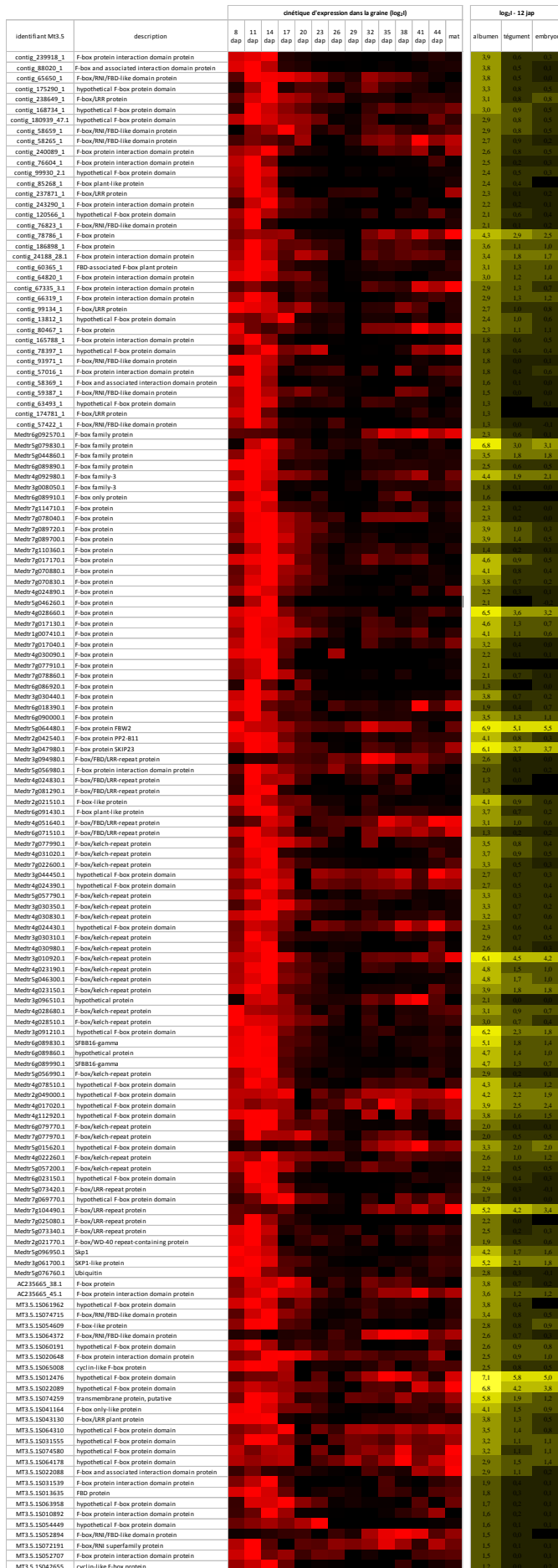


Figure III-10 – Liste des 145 gènes et contigs appartenant à la classe "29.5.11.4.3.2.protein.degradation.ubiquitin. E3.SCF.FBOX " et spécifiquement ou préférentiellement exprimés dans l’albumen. Cinétique d'expression au cours du développement de la graine (données non publiées – K. Righetti, J. Buitink – UMR IRHS Angers) et valeurs des log₂ dans les trois tissus de la graine à 12 jup.

et de la graine, et que leur expression est sensible aux variations environnementales, telles que les stress abiotiques ou la lumière, suggérant une implication dans ces différents processus de développement et/ou de défense (Jain et al., 2007). Chez le blé, les mêmes hypothèses quant au rôle des gènes de type E3 ubiquitine-ligase dans le développement de la graine ont été émises de par les variations de leur profil d'expression tout au long du développement (Capron et al., 2012). De plus, il est suggéré que les gènes impliqués dans les voies de synthèse hormonale et ceux relatifs au complexe Ubiquitin Proteasome System (UPS) agiraient ensemble dans les processus de développement de la graine (Capron et al., 2012). Ces exemples témoignent du lien qui existe entre les voies de dégradation ubiquine/protéasome dépendante, et notamment les protéines de type F-box dans les différentes étapes du développement, la réponse hormonale ou encore l'adaptation face aux fluctuations environnementales, comme décrit par Callis et Vierstra (Callis and Vierstra, 2000).

Les profils d'expression de 145 gènes et contigs de la classe "29.5.11.4.3.2. protein.degradation. ubiquitin.E3.SCF.FBOX " annotés pour la plupart comme des protéines F-box au cours du développement de la graine de 8 jap à maturité (données non publiées – K. Righetti, J. Buitink – UMR IRHS Angers) sont présentés dans la figure III-10. Ce "heatmap" révèle une expression préférentielle de ces gènes, et notamment des gènes codant des protéines de type F-box en début de développement, aux stades 11 et 14 jap, proches du stade de transition étudié. Ceci renforce l'hypothèse du rôle de cette famille de gènes dans le contrôle des divisions cellulaires qui interviennent lors de l'embryogénèse, à la fois dans le développement de l'albumen et de l'embryon. On pourrait également envisager que ces protéines de type F-box marquent les protéines de l'albumen en vue de leur dégradation future, et que ce marquage intervient principalement à cette phase de transition entre embryogénèse et remplissage/maturation, la dégradation ayant ensuite lieu tout au long du développement de la graine jusqu'à la disparition complète de l'albumen lorsque la graine est mature.

c - Classe fonctionnelle "26.miscellaneous enzyme family"

Cette classe qui regroupe divers types d'enzymes est surreprésentée dans le tégument et l'embryon [figure III-7]. Une analyse plus détaillée des codes BIN de cette classe a permis de mettre en évidence une prédominance de la famille de gènes codant les protéines possédant le groupe prosthétique cytochrome P450 dans le tégument et l'embryon [figure III-11]. Cette famille de protéines est très large chez les plantes et est associée à un grand nombre de processus biologiques comme la synthèse des composés de défense, des acides gras ou encore des hormones (Woodward and Bartel, 2005; Mizutani and Ohta, 2010). Parmi la liste de gènes codant des cytochromes P450 dans l'embryon et le tégument, on retrouve un gène (Medtr8g02740.1) présentant une forte homologie de séquence avec le gène EOD3 codant le cytochrome P450/CYP78A6 (e-value Pblast $9,0 \times 10^{-169}$) décrit chez *A. thaliana* comme un régulateur de la taille de la graine via le contrôle de la prolifération des cellules du tégument (Fang et al., 2012). Ce gène Medtr8g02740.1 partage également une forte homologie avec le cytochrome P450 KLUH (KLU)/CYP78A5 chez *A. thaliana*

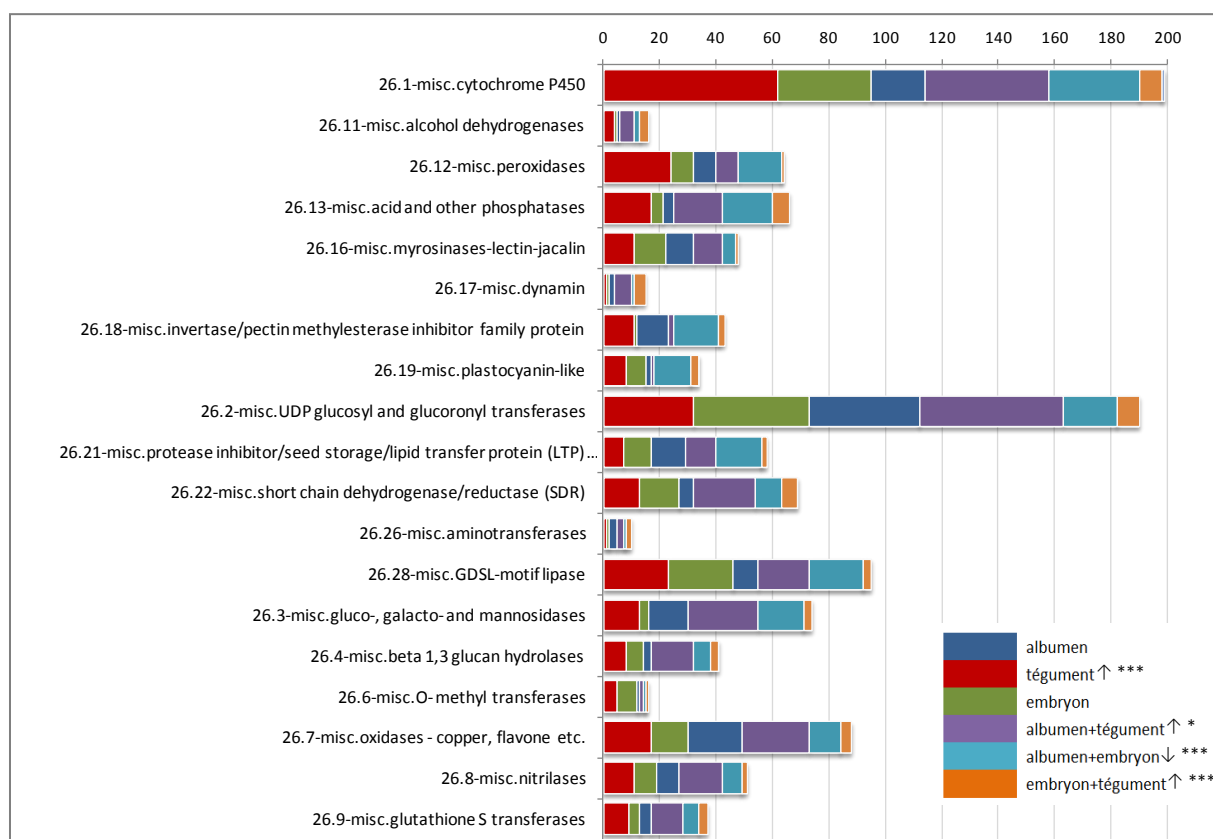


Figure III-11 – Nombre de gènes et contigs de la classe BIN "26.miscellaneous enzyme family" classés selon les différentes sous-familles. La distribution de ces gènes dans les trois tissus (expression spécifique ou préférentielle) est présentée par un code couleur. Les familles comportant moins de 10 gènes ne sont représentées sur cette figure. Les symboles ↑ et ↓ correspondent respectivement aux familles sur- et sous-représentées dans le(s) tissu(s) d'après la figure III-7 et l'annexe III-7 avec des valeurs p de probabilité : *** $p < 0,001$; ** $p < 0,01$, * $p < 0,05$.

(e-value Pblast $6,9^{e-123}$). Chez *A. thaliana*, ce gène est spécifiquement exprimé dans les couches internes du tégument de la graine régulant ainsi sa prolifération cellulaire et par conséquent intervenant dans le contrôle de la taille de la graine (Adamski et al., 2009). Chez le riz, un gène de la famille CYP78A a été décrit comme impliqué dans le développement de l'embryon. Le phénotype "embryon géant" serait lié à la sur-expression d'un gène GE (Giant Embryon) codant le cytochrome CYP78A13 et dont l'expression dans la graine est restreinte aux couches cellulaires de l'albumen et aux cellules de l'embryon qui se font face. Les deux sites d'expression seraient nécessaires à ce gène pour contrôler le développement coordonné de l'embryon et de l'albumen (Nagasawa et al., 2013; Yang et al., 2013)

Parmi les sous-familles présentées dans la figure III-11, on retrouve une large part d'enzymes de type "UDP glucosyl and glucoronyl transferases"; ces glycosyl-transférases assurent le transfert de groupement monosaccharide sur des composées aglycones, tels que les hormones ou les métabolites secondaires et composées intervenant dans les processus de défense et de tolérance aux stress (Vogt and Jones, 2000; Luang et al., 2013). Des séquences codant des UDP-glycosyltransférases ont été identifiées par RT-PCR dans des graines de colza en cours de développement et malgré des profils d'expression variés, une abondance de transcrits a été observée dans les graines jeunes (Mittasch et al., 2007).

Des enzymes de type GDSL-motif lipase sont également présentes en nombre dans l'embryon et le tégument [figure III-11]. Ces enzymes clivent les mono-, di- et tri- acylglycérols en acide gras et alcool et sont requises dans les processus de défense contre les pathogènes et lors des stress environnementaux (Hong et al., 2008). Par ailleurs, l'étude du transcriptome chez des plantes présentant la mutation *ZOU* ou *RGE* (Yang et al., 2008; Kondou et al., 2008), qui induit la formation de petites graines sèches et déformées, a montré que des gènes de types GDSL-lipases étaient sous-exprimés chez ce mutant, de même qu'un gène codant un cytochrome P450. Les GDSL-motif lipases pourraient donc aussi intervenir dans le contrôle du développement de la graine (Kondou et al., 2008). Il est important de noter que l'on retrouve ces familles de gènes majoritairement sous-exprimés dans l'analyse comparative du transcriptome de gousses *dash* EMS109 vs WT à 8 jap (cf tableau supplémentaire 1, article en préparation, chapitre 2) suggérant leur possible implication dans le contrôle de la taille de la graine.

III.2.4 - Analyse des gènes exprimés pendant l'embryogénèse

Afin de déterminer quels sont les gènes préférentiellement exprimés pendant les phases d'embryogénèse et de transition par rapport aux autres stades, nous avons utilisés des données transcriptomiques générées pendant le développement de la graine de l'embryogénèse (8jap) jusqu'à maturité (44jap) (données non publiées – K. Righetti, J. Buitink – UMR IRHS Angers). Les données d'expression relatives aux stades embryogénèse et transition (soit 8, 11 et 14 jap) ont été moyennées, de même que celles relatives aux stades maturation (35, 38, 41 et 44 jap) et le ratio de ces deux moyennes a été calculé afin de mettre en évidence les gènes les plus fortement exprimés en début de développement

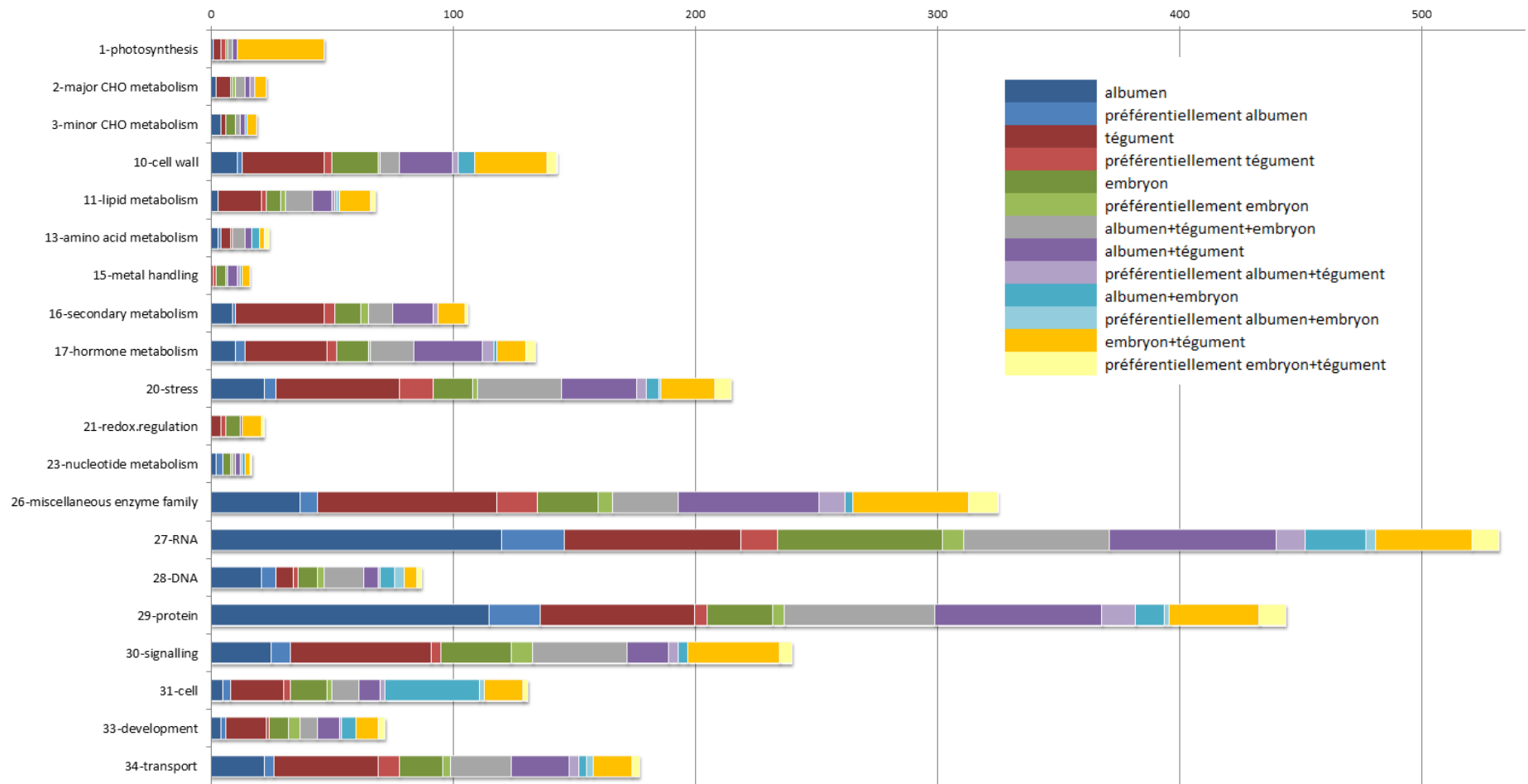


Figure III-12 – Analyse des gènes préférentiellement exprimés au cours de l'embryogénèse (8-11 j). Répartition du nombre de gènes par classe fonctionnelle et par spécificité ou préférence tissulaire. Les classes pour lesquelles le nombre de gènes est inférieur à 10 ne sont pas représentées. Les 726 gènes de la classe fonctionnelle « 35.not assigned » n'ont pas été pris en compte.

de la graine. Cette liste de gènes a ensuite été croisée avec les 47 691 gènes et contigs exprimés dans les tissus de la graine à 12 jap (cf tri n°1-tableau III-1) et nous a permis de mettre en évidence 4 614 gènes ayant une expression cinq fois plus importante au début du développement de la graine. La répartition de ces gènes en fonction des classes fonctionnelles BIN et selon les tissus dans lesquels ils sont spécifiquement ou préférentiellement exprimés est présentée dans la figure III-12. Les classes pour lesquelles le nombre de gènes est inférieur à dix ne sont pas représentées, de même que les 726 gènes de la classe fonctionnelle "35.not assigned".

Il est intéressant de noter que les classes fonctionnelles "27.RNA", "29.protein" et "26.miscellaneous enzyme family" sont les plus représentées à des stades précoces du développement de la graine. Ces classes ressortaient déjà parmi celles les plus fortement représentées dans les analyses ciblées des gènes exprimés spécifiquement ou préférentiellement dans les tissus (cf. figure III-6 et III-7), et confortent donc les interprétations énoncées jusqu'à présent. De façon surprenante mais déjà suggéré dans des travaux antérieurs (Loftfield et al., 1953; Palma et al., 2002) et lors de l'analyse précédente de la classe fonctionnelle "29.protein", la majorité des gènes de cette classe sont liés à la dégradation et non à la synthèse (cf paragraphe III.2.3.2.b - Classe fonctionnelle "29.protein"). On retrouve en effet de nombreux gènes codant des protéases à cystéine, protéases à aspartate ou protéases à sérine ainsi que des subtilases, ou encore des gènes impliqués dans les voies de dégradation *via* l'ubiquitine. Les fonctions associées à ces protéases ne sont pas toutes liées à la dégradation ; elles interviennent en effet dans la séparation des protéines associées à des signaux d'adressage ou encore dans l'approvisionnement en acides aminés pour la synthèse de nouvelles protéines (Vierstra, 1996). Certaines subtilases ont également été décrites comme fortement exprimées dans les tissus de la graine suggérant un rôle associé à son développement (Batchelor et al., 2000; Yang et al., 2008; D'Erfurth et al., 2012). La répartition tissulaire de cette famille fonctionnelle indique une forte proportion de gènes exprimés dans les trois tissus, mais également un nombre important de gènes spécifiques ou préférentiellement exprimés dans le tégument, certaines subtilases impliquées dans le développement du tégument ont ce même profil d'expression (*ie* spécificité d'expression dans le tégument) (Batchelor et al., 2000). On note d'autres spécificités tissulaires comme une forte proportion de gènes exprimés dans l'embryon et le tégument (expression spécifique ou préférentielle) dans la classe "1.photosynthesis". Les gènes spécifiques ou préférentiellement exprimés dans l'albumen se retrouvent de façon majoritaire dans la classe fonctionnelle "27.RNA", ce qui est cohérent avec le développement et la cellularisation importante de ce tissu à ce stade d'embryogénèse.

III.2.5 - Analyse des gènes impliqués dans le métabolisme de l'AIA

Il nous a semblé important d'analyser de façon détaillée la répartition tissulaire des gènes impliqués dans le métabolisme hormonal. Ainsi les gènes relatifs aux sous-classes fonctionnelles BIN "17.2.hormone metabolism.auxin", "27.3.4.RNA.regulation of transcription.ARF, Auxin Response Factor family" et

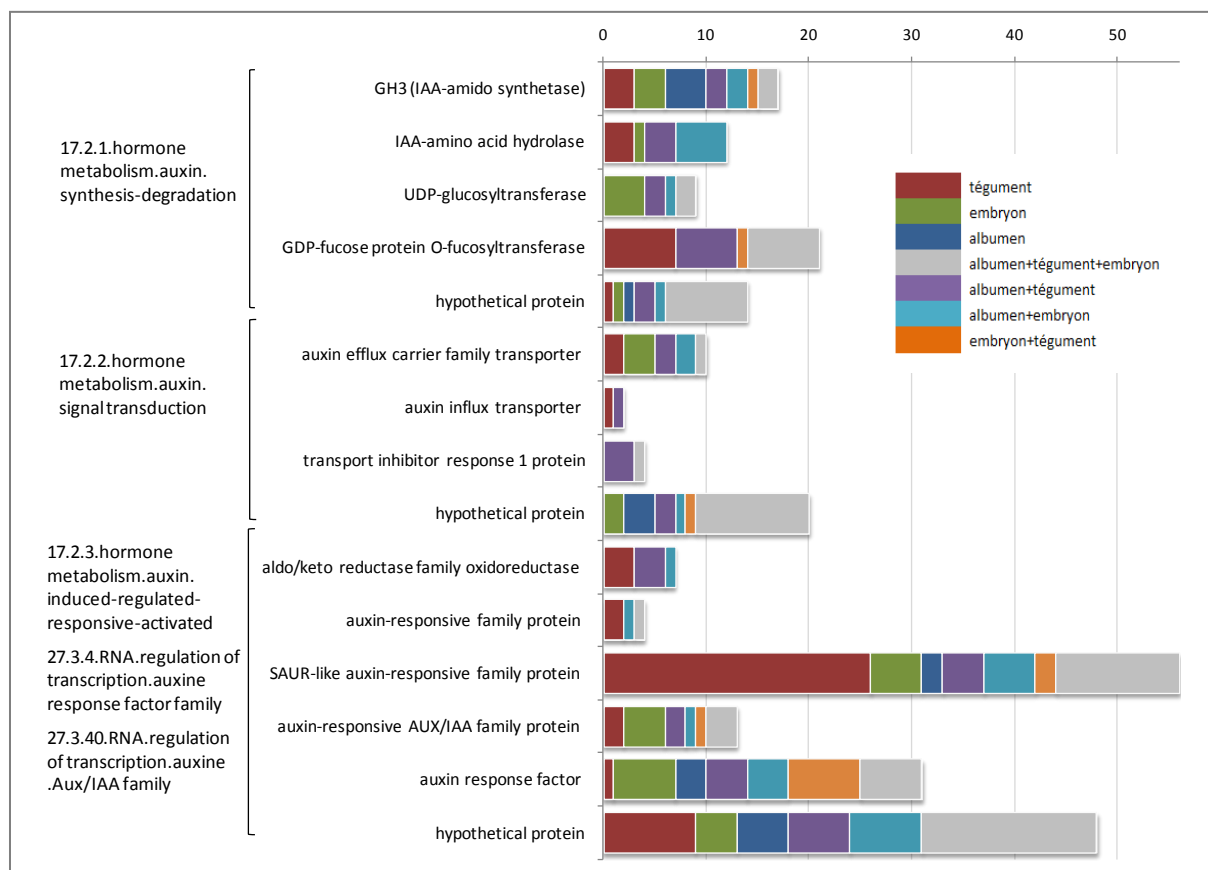


Figure III-13 – Nombre de gènes et contigs impliqués dans le métabolisme de l'auxine (sous classes BIN "17.2.hormone metabolism.auxin", "27.3.4. RNA.regulation of transcription.ARF, Auxin Response Factor family" et "27.3.40. RNA.regulation of transcription.Aux/IAA family"). La distribution de ces gènes dans les trois tissus (expression spécifique et préférentielle) est présentée par un code couleur.

"27.3.40.RNA.regulation of transcription.Aux/IAA family" ont été sélectionnés. Les annotations relatives aux sous-classes fonctionnelles BIN n'étant pas très détaillées, il nous a paru plus judicieux de rechercher les annotations relatives aux identifiants Mt3.5 et Mt4.0 de ces gènes afin d'avoir une analyse plus ciblée des familles de gènes représentées. La distribution de ces familles de gènes dans les trois tissus (expression spécifique et préférentielle) est présentée dans la figure III-13.

Parmi les gènes relatifs à la synthèse ou dégradation de l'AIA, on retrouve majoritairement des enzymes impliquées dans les conversions de l'AIA avec ses formes conjuguées. Jusqu'à présent, aucune étude n'a permis de définir le rôle de ces formes conjuguées dans le développement de la graine, mais seulement de mettre en évidence leur importance dans le stockage de l'AIA et d'identifier certains composés (AIA-Asp et AIA-Glu) conduisant à la dégradation de l'AIA (Ludwig-Muller, 2011). Les enzymes GH3 de conjugaison de l'AIA avec les acides aminées ainsi que les enzymes liées à l'hydrolyse de ces conjugués (AIA-amino acid hydrolase) sont exprimées dans les trois tissus. On observe des spécificités tissulaires plus marquées concernant les enzymes de conversions de ces conjugués en dérivés glycosylés (conjugués d'esters). Ainsi, les UDP-glucosyltransférases semblent plus spécifiques de l'embryon, alors que les GDP-fucose-O-fucosyltransférases se retrouvent à la fois dans l'albumen et le tégument. Il est à noter que les UDP-glucosyltransférases sont également retrouvées dans la classe fonctionnelle "26.2.misc.UDP glucosyl and glucoronyl transferases" et ont été précédemment décrites comme fortement représentées dans les tissus (cf paragraphe III.2.3.2.c). Quant aux gènes relatifs au transport, on observe une majorité de transporteurs d'efflux par rapport au transport d'influx. Les transcrits de ces transporteurs sont présents dans tous les tissus, excepté dans l'albumen. On observe cependant une part importante de gènes communs à deux tissus (*ie* embryon et albumen ou tégument et albumen) suggérant l'existence de transporteurs d'auxine spécifiques entre ces tissus. L'analyse des gènes de réponses à l'AIA indique une forte proportion de gènes *SAURs* dans le tégument, la fonction de ces gènes reste indéterminé à ce jour bien que quelques *SAURs* aient déjà été décrits dans la littérature comme impliqués dans la régulation négative de la synthèse d'AIA et de son transport (*SAUR39* (Kant et al., 2009; Kant and Rothstein, 2009)). Les autres gènes de réponse à l'AIA de type Aux/IAA ou ARF se retrouvent exprimés dans les trois tissus avec une dominance dans l'embryon.

Lorsque l'on s'intéresse plus particulièrement aux familles de gènes retrouvées dans l'albumen (*DASH* étant fortement exprimé dans ce tissu, chapitre II), on note une prédominance des gènes GH3 et des FTs de type ARF. Les gènes ARF ont précédemment été décrit comme régulant négativement les divisions cellulaires, contrôlant ainsi un développement "normal" des graines (Schruff et al., 2006). Parmi les cinq gènes ARF dérégulés d'après l'analyse comparative du transcriptome de gousses *dash* EMS109 vs WT à 8 jap (cf tableau supplémentaire 1, article en préparation, chapitre2), un seul est présent dans l'albumen à 12 jap (Medtr1g094960.1). Le gène *ARF16* correspondant à son orthologue chez *A. thaliana*, semblerait impliqué dans le contrôle de la mise en place de la dormance dans les graines (Liu et al., 2013).

III.3 - Conclusion

Durant les premiers stades de développement de la graine, alors que le développement de l'embryon est basé sur des divisions et différenciations cellulaires, l'albumen quant à lui se développe d'abord en un syncytium qui devient progressivement un tissu cellularisé (cf. figures II-5 et annexe II-3 - cytologie du développement de la graine - article en préparation, chapitre 2). Cette évolution de structure de l'albumen est étroitement liée au développement de l'embryon tout au long du développement de la graine puisqu'il deviendra une réserve de nutriments nécessaires à l'élongation cellulaire de l'embryon. A 12 jap, phase de transition entre l'embryogénèse et le remplissage, de nombreux changements interviennent et notamment dans le développement de l'embryon qui devient sensible aux signaux hormonaux et aux composés nutritifs et acquiert ses capacités de stockage (Weber et al., 2005). L'étude du transcriptome des tissus de la graine à 12 jap visait à mettre en évidence des voies métaboliques liées à ces changements ou des spécificités tissulaires susceptibles d'interpréter ce stade de transition. L'analyse des classes fonctionnelles révèle une forte proportion de gènes appartenant aux classes "27.RNA", "29.protein" et "26.misc"; les classes "27.RNA" et "29.protein" ayant déjà été décrites comme fortement représentées d'après l'étude du protéome nucléaire de la graine à 12 jap (Repetto et al., 2008). A travers cette étude, la régulation du transcriptome par les FTs apparaît une nouvelle fois déterminante dans le contrôle et la mise en place des processus conduisant au développement de la graine. En effet, certaines familles de FTs ont été identifiées comme préférentiellement exprimés dans un tissu [figure III-2 et III-8] : FTs de type B3 dans l'embryon, FTs associés au métabolisme hormonal Aux/IAA dans le tégument ou les WRKY dans l'albumen. Il est toutefois difficile d'assigner une spécificité d'expression tissulaire à certaines familles de FTs en ne ciblant qu'un seul stade de développement, comme souligné dans l'étude des FTs de la graine chez *A. thaliana* (Le et al., 2007). Plusieurs gènes d'une même classe fonctionnelle peuvent agir en synergie tout en ayant des profils d'expression différents, ce qui ne peut pas être pris en compte avec cette analyse du transcriptome à un stade isolé.

Cette étude a tout de même permis d'identifier des spécificités tissulaires. Ainsi, ces analyses ont permis de souligner l'importante part de gènes liés à la dégradation via l'ubiquitine dans l'albumen, suggérant ainsi la mise en place dès le stade transitoire du mécanisme de mort cellulaire de l'albumen, qui disparaît progressivement au profit du développement de l'embryon (cf introduction). Ce développement coordonné de l'albumen et de l'embryon est également mis en évidence de par les nombreuses activités cellulaires liées à la transcription, ce qui laisse supposer que la transcription de l'embryon serait plus liée à son expansion cellulaire et à sa croissance alors que la transcription dans l'albumen serait plutôt vouée à la synthèse de protéines de dégradation.

III.4 - Matériels et méthodes

III.4.1 - Plan d'expérience

La technique du dye-switch (Churchill, 2002) a été retenue pour cette expérience. Ce principe repose sur une hybridation en bi-couleurs de deux échantillons marqués sur une même puce. Les couleurs sont inversées lors des répétitions biologiques afin d'éliminer les biais de marquage. Le plan de marquage retenu pour l'hybridation des trois tissus est le suivant:

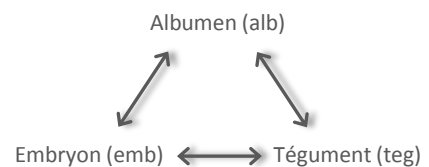
emb1-Cy5 versus teg3-Cy3	alb1-Cy3 versus teg1-Cy5	emb1-Cy5 versus alb1-Cy3
emb2-Cy3 versus teg1-Cy5	alb2-Cy5 versus teg2-Cy3	emb3-Cy3 versus alb2-Cy5
emb3-Cy5 versus teg2-Cy3	alb3-Cy3 versus teg3-Cy5	emb2-Cy5 versus alb3-Cy3

avec l'échantillon teg3 marqué une fois avec Cy3 et une autre fois avec Cy5

teg pour tégument, alb pour albumen et emb pour embryon

1, 2 et 3 étant les répétitions biologiques

L'ensemble des comparaisons possibles est représenté dans le triangle ci-contre avec 3 hybridations pour chaque répétition.



III.4.2 - Obtention du matériel biologique - préparation des échantillons

Les plantes utilisées, de génotype A17 ont été cultivées sous serre dans des pots de 1,5L rempli d'un mélange attapulgite : billes d'argiles (60 :40). Les conditions de température et de luminosité étaient régulées : 20°C le jour avec une photopériode de 16 heures et 18°C la nuit. L'apport en azote n'était pas limitant tout au long du développement des plantes, avec un apport 3,5N/3,1P/8,6K.

Les fleurs ont été notées et les gousses collectées à 12 jap selon Gallardo et al. (Gallardo et al., 2007). Les graines ont été disséquées sous loupe binoculaire dans des boîtes de Pétri placées sur de la glace, les tissus ont été prélevés à partir de ces graines et immédiatement congelés dans l'azote liquide avant d'être stockés à -80°C.

III.4.3 - Extraction d'ARN et obtention des ARNa (ARN amplifiés)*

Les tissus ont été broyés en tube Eppendorf 2,0mL avec une bille de verre grâce à un broyeur de type Retsh et le tampon d'extraction ajouté immédiatement après le broyage. L'extraction des ARNs totaux a été réalisée grâce au kit RNeasy Micro Kit (Quiagen) selon les instructions du fabricant. L'obtention des ARNa a été soumise à plusieurs étapes : transcription inverse à l'aide du kit Message AmpII a ARN Amplification (Ambion), purification des ADNc ainsi obtenus grâce au kit DNAClear (Ambion), transcription in vitro et purification des ARNa en utilisant le kit MEGAClear (Ambion). Un ajustement au ¼ a été fait pour l'ensemble de ces kits afin de réduire le volume utilisé et de maximiser le nombre d'échantillons par kit. Les

ARNa ainsi obtenus ont ensuite été quantifiés au NanoDrop et concentrés par précipitation à -20°C après ajout d'acétate de sodium 3M (1/10 de volume) et de 2,5 volume d'éthanol absolu en présence d'acrylamide (1/100 de volume). Après lavage, le culot est repris dans de l'H₂O DEPC pour une concentration d'environ 1 µg.L⁻¹.

* L'obtention des ARNa ainsi que l'ensemble des manipulations jusqu'à l'hybridation des puces ont été réalisés au laboratoire IRHS à Angers, sous la responsabilité de Julia BUITINK et de Sandra PELLETIER qui a également réalisé les analyses d'images et les analyses statistiques sous R (scripts permettant la normalisation et l'analyse différentielle) afin d'obtenir des fichiers de sorties sous excel.

III.4.4 - Marquage des ARNa

Le marquage des ARNa est réalisé par transcription inverse en présence de dCTP marqués au Cy3/ Cy5 (cyanine rouge/ verte). Une première étape consiste à incuber dans un volume final de 10,5 µL un mélange de 5 µg d'ARNa, 2 µg de Random nonamères (Sigma) et 20 U d'inhibiteur de RNase (Ribolock, Thermo Scientific) à 65°C pendant 10 min. Après ajout de 20 µL d'un mix marquage (composé de tampon « first strand » (Invitrogen) 1X final, DTT 10 mM final (Invitrogen), dDTP 500 µM final, dCTP 100 µM final, Cy5 ou Cy3 selon l'échantillon 75 µM final (Interchim) et 200 U de Superscript revertase III (Invitrogen)), ce nouveau mélange est incubé 2h à 42°C. 200 U de Superscript revertase III (Invitrogen) sont ajoutés au bout d'1 h d'incubation. Après ajout de 79 µL d'H₂O DEPC, les ADNc ainsi marqués sont purifiés à l'aide du kit Nucleospin PCR Clean-up (Macherey-Nagel) en suivant le protocole dispensé par le fournisseur, l'élution finale est réalisée avec 2 x 21 µL d'H₂O DEPC préchauffée à 55°C. La concentration en ADNc (ng/µL) et en Cy5 (ou Cy3) (pmol/µL) est ensuite mesurée à l'aide du NanoDrop. Les mélanges d'échantillons marqués sont réalisés en suivant le plan d'expérience : 30 pmol d'ADNc marqué Cy3 et 30 pmol d'ADNc marqué Cy5 et précipités à -20°C après ajout d'acétate de sodium 3M (1/10 de volume) et de 2,5 volume d'éthanol absolu en présence d'acrylamide (1/100 de volume). Les culots secs sont ensuite conservés à -20°C à l'obscurité en attendant l'étape d'hybridation.

III.4.5 - Hybridation des puces et scan des lames

Les culots contenant les ADNc marqués des deux échantillons hybridés sur la puce sont resuspendus dans 2,2 µL de STC (Sample Tracking Control) (NimbleGen). 5,8 µL d'un mix de solution d'hybridation (tubes 1, 2 et 3 du kit d'hybridation (NimbleGen)) sont ensuite ajoutés avant une incubation de 5 min à 95°C. Une fois le montage de la lame réalisée (fixation du mixer sur la lame à l'aide du PMAT), celle-ci est placée à 42°C pendant 5 min puis 6,3 µL d'échantillon sont ensuite déposés sur chacune des puces grâce au mixer collé sur la lame. Cette dernière est placée 16-20h à 42°C dans la station d'hybridation puis le mixer est décollé de la lame que l'on introduit successivement dans plusieurs solutions de lavage : 2 min dans un

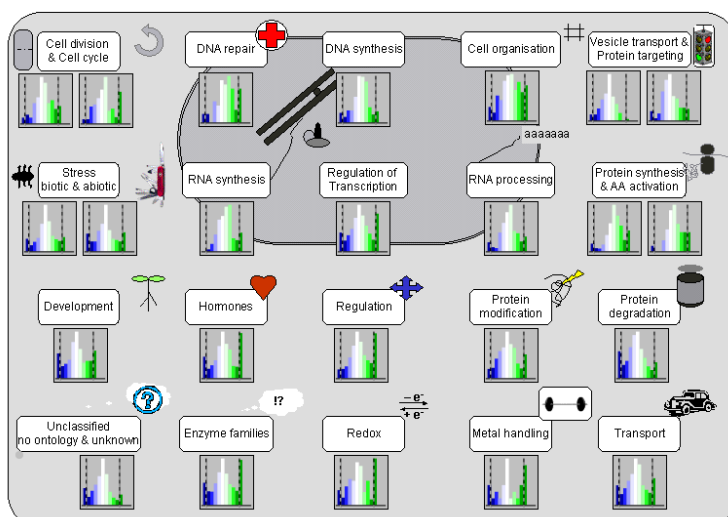
tampon de lavage I (NimbleGen) supplémenté de DTT 100 μ M et préchauffé à 42°C, 1 min dans le tampon de lavage II (NimbleGen) supplémenté de DTT 100 μ M et enfin 15 sec dans le tampon lavage III (NimbleGen) supplémenté de DTT 100 μ M. La lecture des lames 12-plex NimbleGen est réalisée par le scanner MS 200 (NimbleGen) via le logiciel MS 200 v1.2 qui génère 2 images : marquage vert (532 nm) et marquage rouge (635 nm).

III.4.6 - Analyse des données, traitement et conversion

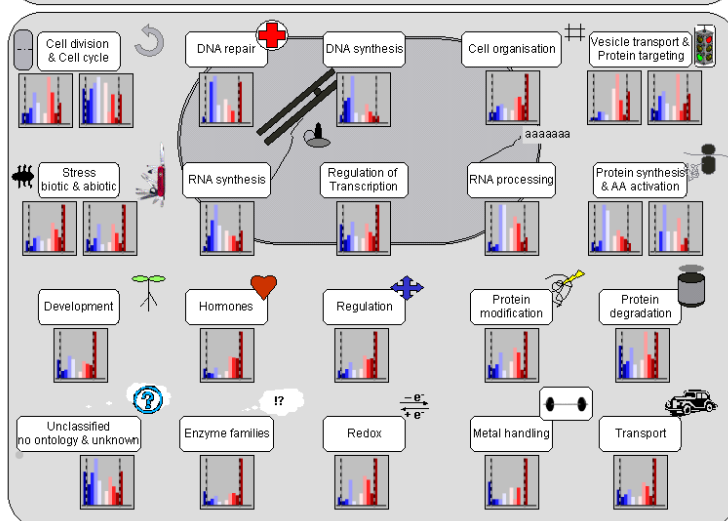
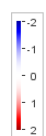
Les analyses d'images sont réalisées par le logiciel DEVA v1.2.1, les signaux identifiés par alignement de la grille d'identification des sondes sont convertis en valeurs numériques pour chaque couleur. Les fichiers de sorties sont ensuite analysés via le langage R (R v2.14 : <http://www.rproject.org>, Rstudio : <http://www.rstudio.com>) et le script "pair" (développé par le laboratoire IRHS à Angers). Les données ont ensuite été normalisées par la méthode lowess, puis les \log_2 des intensités et des ratios (respectivement $\log_2 I$ et $\log_2 RI$) calculés et moyennés via le package "LIMMA" (Linear Models for Micro Array data, Bioconductor project (<http://www.bioconductor.org>) (Smyth et al., 2005)) avec un test Bayésien pour définir les sondes significativement différentiellement exprimées. L'ensemble des analyses statistiques de normalisation des données, l'analyse différentielle et le contrôle des faux-positifs ont été réalisé par Sandra PELLETIER (UMR IRHS, Angers). La conversion et le traitement des données permettent d'obtenir une sortie .txt ou .xlsx avec les noms des sondes associés à leurs valeurs p de probabilité, valeurs d'intensité et mesures de ratio. L'ensemble des valeurs d'intensités brutes (I) est exprimé en $\log_2 I$, normalisé par la méthode de Lowess (régression linéaire par intervalle sans soustraction du bruit de fond). Les couples ratio/intensités sont soumis à un test Bayésien qui leur attribue une valeur p de probabilité et le contrôle des faux-positifs est réalisé par un test FDR (False Discovery Rate). Enfin une correction des valeurs p de probabilité est utilisée afin d'attribuer une valeur BH (Benjamini Hochtberg) suivant l'indice de confiance qu'il lui porte. Les données de ces puces NimbleGen sont disponibles sur le portail GEO Gene Expression Omnibus (GPL16373).

III.5 - Annexes

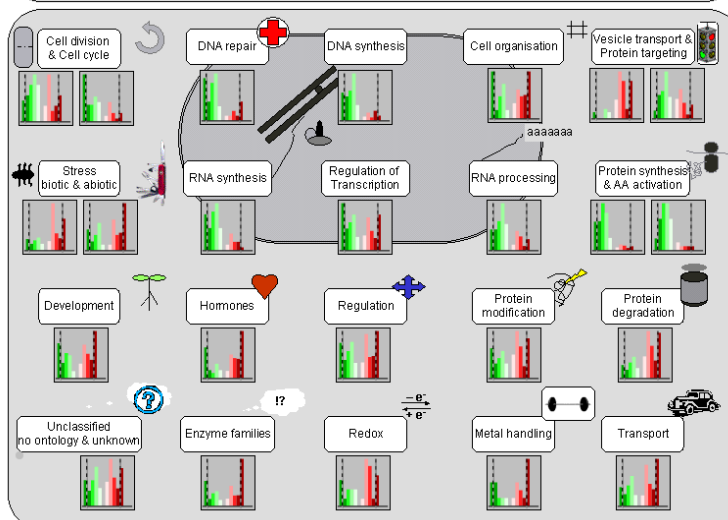
A
embryon/albumen
21733 gènes



B
tégument/albumen
11021 gènes

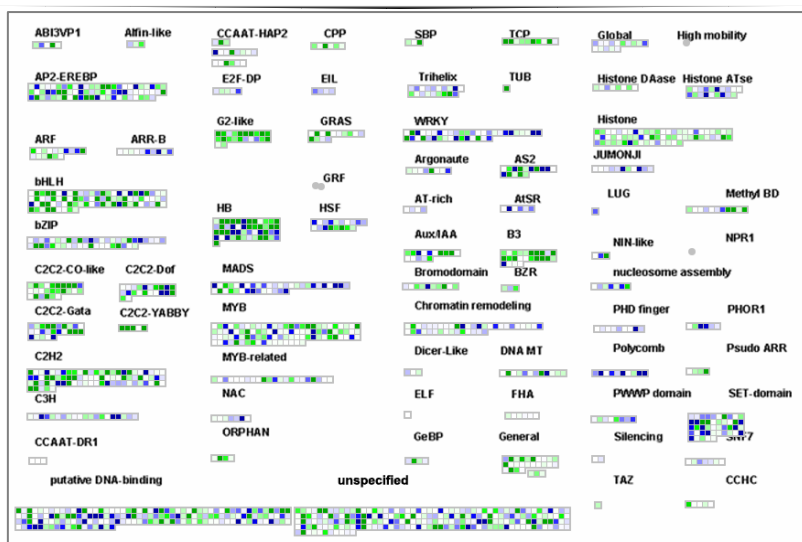


C
tégument/embryon
10659 gènes

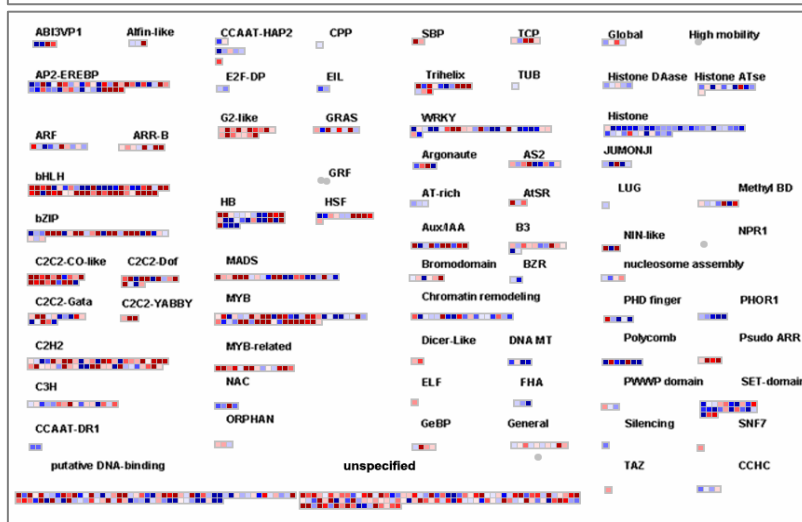
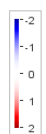


Annexe III-1 – Analyse comparative des fonctions cellulaires. A. albumen/embryon. B. albumen/tégument. C. embryon/tégument. Les hauteurs des histogrammes sont fonctions du nombre de gènes et les intensités des couleurs sont relatives à l'expression plus ou moins fortes des gènes dans les tissus : vert : embryon ; bleu : albumen ; rouge : tégument. Pour chaque comparaison, le nombre de gènes présents sur la cartographie correspondant à ceux pour lesquels un code BIN a pu être attribué est indiqué.

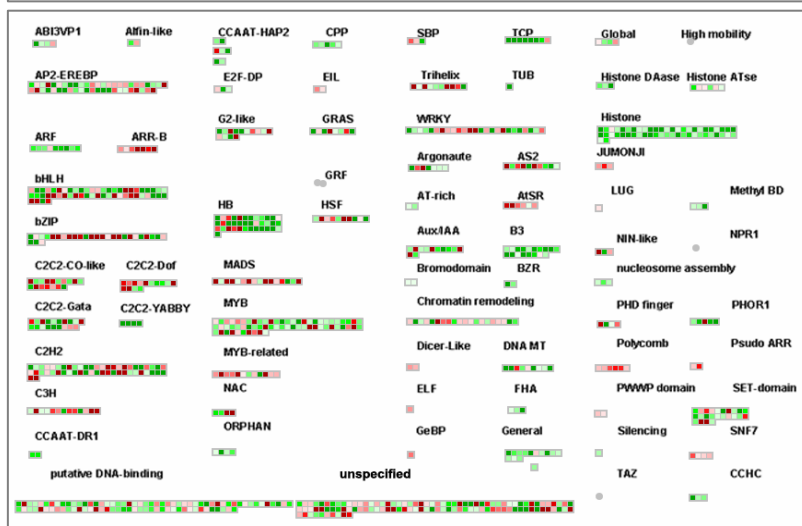
A
embryon/albumen
1591 gènes



B
tégument/albumen
902 gènes



C
tégument/embryon
944 gènes



Annexe III-2 – Analyse comparative des cartes "transcription" générées par le logiciel MapMan. Chaque carré de couleur représente un gène et les intensités des couleurs sont relatives à l'expression plus ou moins forte des gènes dans les tissus : vert : embryon ; bleu : albumen ; rouge : tégument. Pour chaque comparaison, le nombre de gènes présents sur la carte, correspondant à ceux pour lesquels un code BIN a pu être attribué, est indiqué.

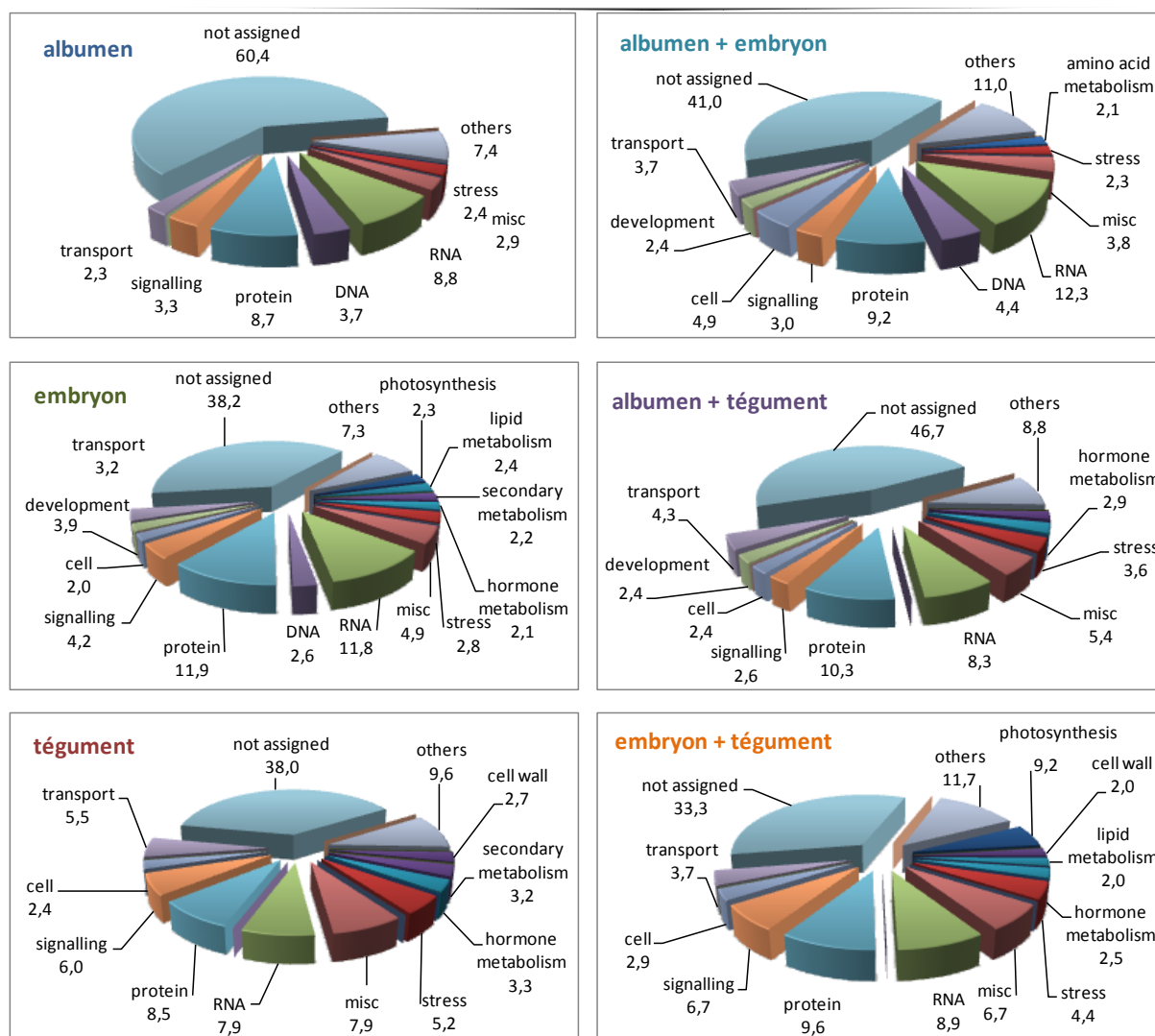
prélèvement	Tissu	masse de tissu (mg)	nombre de graines décortiquées	masse de tissu par graine (mg)	masse d'une graine (mg)	% de tissu pour 1 graine
n°1	tégument	164,4	107	1,54	2,7	57,0
	albumen	48,3		0,45		16,7
	embryon	75,8		0,71		26,3
n°2	tégument	186,9	126	1,48	2,9	50,9
	albumen	63,5		0,50		17,3
	embryon	116,6		0,93		31,8
n°3	tégument	125,8	138	0,91	1,8	50,7
	albumen	50,9		0,37		20,5
	embryon	71,5		0,52		28,8
n°4	tégument	68,1	50	1,36	2,7	50,0
	albumen	28,3		0,57		20,8
	embryon	39,8		0,80		29,2
n°5	tégument	96,5	56	1,72	3,3	51,9
	albumen	37,0		0,66		19,9
	embryon	52,5		0,94		28,2
n°6	tégument	99,4	54	1,84	3,5	52,4
	albumen	39,4		0,73		20,8
	embryon	50,7		0,94		26,8
n°7	tégument	105,6	57	1,85	3,6	51,9
	albumen	41,3		0,72		20,3
	embryon	56,3		0,99		27,7

soit au total pour 1 graine, le **tégument** représente **52,15% ± 0,94**

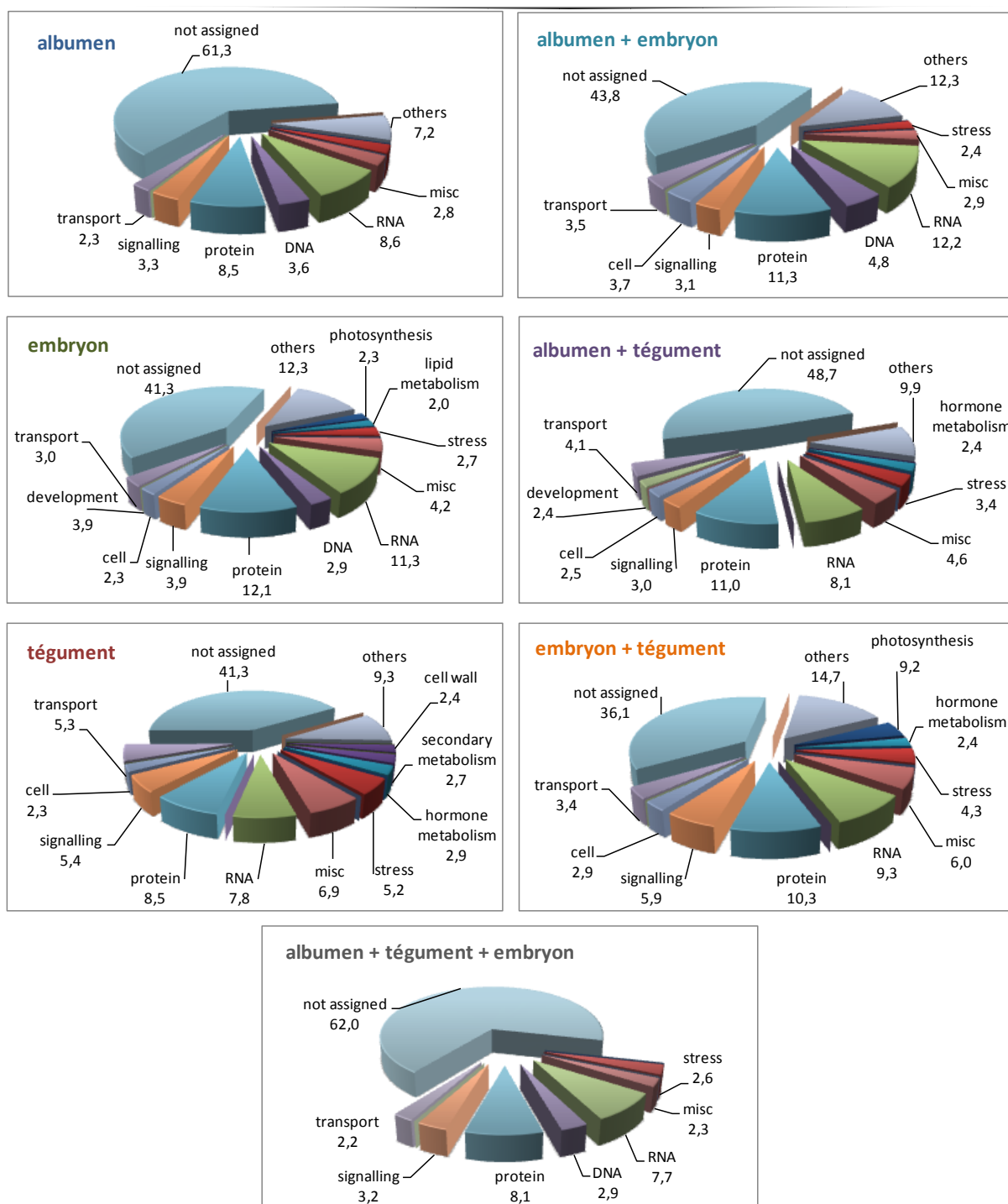
l'**albumen** représente **19,34% ± 0,70**

l'**embryon** représente **28,51% ± 0,74**

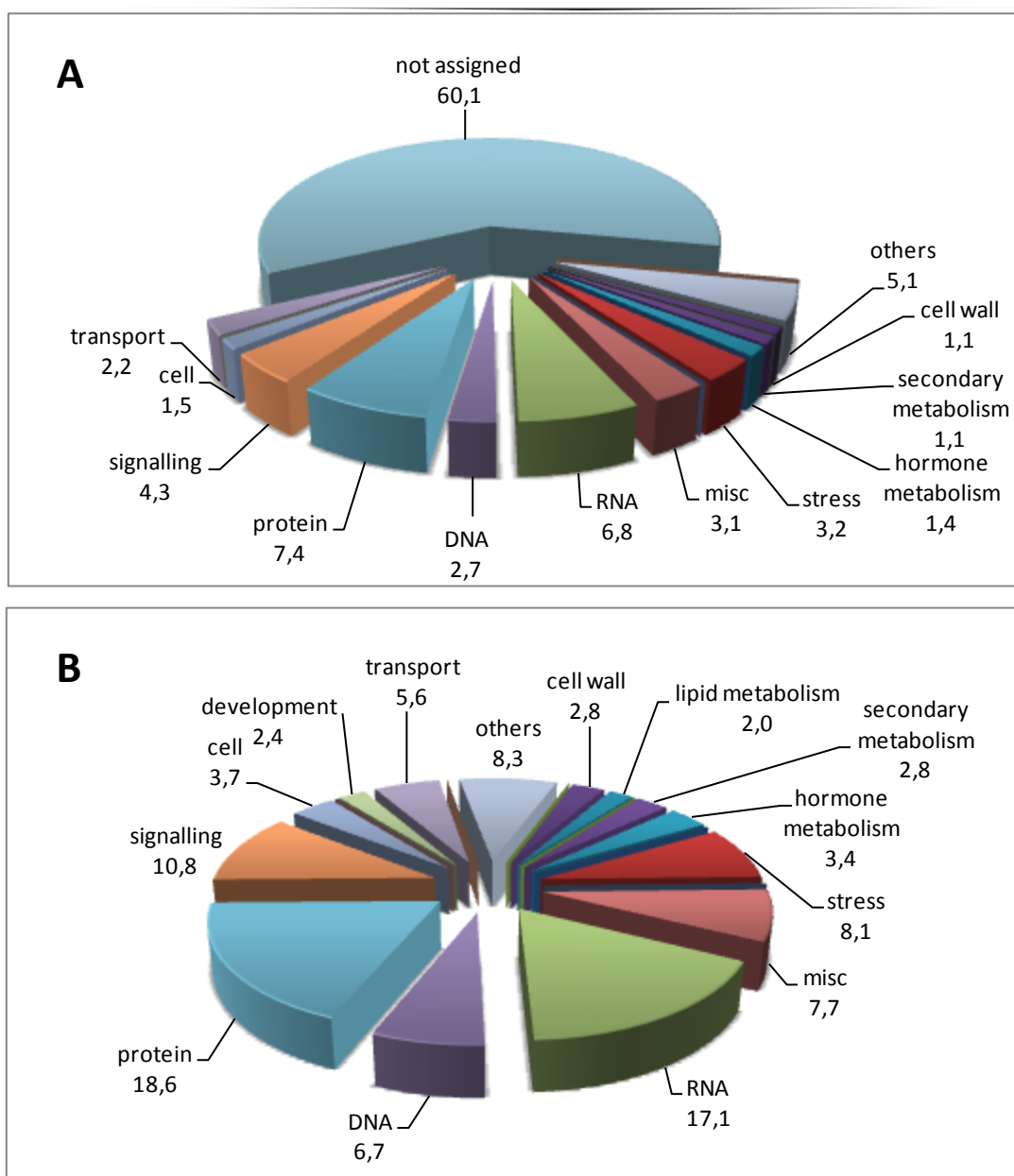
Annexe III-3 – Répartition en mg des trois tissus de la graine lors des préparations d'échantillons en vue d'en déduire un pourcentage relatif à la représentation du tissu au sein de la graine. Tableau descriptif des données présentées dans la figure III-5.



Annexe III-4 – Classification fonctionnelles (BIN code) des gènes spécifiques de chaque tissu : tegument, embryon et albumen et des gènes spécifiquement exprimés dans deux tissus : embryon et tegument, albumen et embryon et albumen et tegument. Les classes dont les pourcentages sont inférieurs à 2% ont été regroupées dans la classe "others".



Annexe III-5 – Classification fonctionnelles (BIN code) des gènes spécifiques et préférentiels de chaque tissu : tegument, embryo et albumen et des gènes spécifiquement ou préférentiellement exprimés dans deux tissus : embryo et tegument, albumen et embryo et albumen et tegument ou dans les trois tissus. Les classes dont les pourcentages sont inférieurs à 2% ont été regroupées dans la classe "others".



Annexe III-6 – A. Classification fonctionnelles (BIN code) des 93445 gènes et contigs présents sur la puce NimbleGen Medtr_v1.0. Les classes dont les pourcentages sont inférieurs à 0,3% ont été regroupées dans la classe "others". B. Classification fonctionnelles (BIN code) des 93445 gènes et contigs présents sur la puce NimbleGen Medtr_v1.0. Ces diagrammes ne font pas apparaître la classe fonctionnelle "35.not assigned", qui n'a pas non plus été prise en compte pour le calcul des pourcentages de représentativité. Les classes dont les pourcentages sont inférieurs à 2% ont été regroupées dans la classe "others".




CHAPITRE -3- Annexes

BinCode-BinName	ensemble des sondes la puce NimbleGen Medtr_v1.0.		analyse des sondes spécifiquement et préférentiellement exprimées dans l'albumen		estimation de la dérégulation d'une classe (enrichissement ou appauvrissement en gènes)	valeur p de probabilité associée à la dérégulation de la classe	analyse des sondes spécifiquement et préférentiellement exprimées dans l'embryon		estimation de la dérégulation d'une classe (enrichissement ou appauvrissement en gènes)	valeur p de probabilité associée à la dérégulation de la classe	analyse des sondes spécifiquement et préférentiellement exprimées dans le tégument		estimation de la dérégulation d'une classe (enrichissement ou appauvrissement en gènes)	valeur p de probabilité associée à la dérégulation de la classe	analyse des sondes exprimées dans les trois tissus		estimation de la dérégulation d'une classe (enrichissement ou appauvrissement en gènes)	valeur p de probabilité associée à la dérégulation de la classe
	nombre de sondes	% répartition	nombre de sondes	% répartition			nombre de sondes	% répartition			nombre de sondes	% répartition			nombre de sondes	% répartition		
1-photosynthesis	548	1,5	10	0,4	-1,1	8,41E-07	102	3,5	2,0	2,19E-13	14	0,6	-0,9	1,13E-04	58	0,8	-0,7	1,57E-07
2-major CHO metabolism	241	0,6	11	0,4	-0,2	2,92E-01	15	0,5	-0,1	4,68E-01	22	0,9	0,3	1,20E-01	74	1,0	0,3	3,30E-03
3-minor CHO metabolism	246	0,7	21	0,9	0,2	2,50E-01	18	0,6	0,0	9,05E-01	18	0,7	0,1	6,04E-01	46	0,6	-0,1	5,86E-01
4-glycolysis	119	0,3	8	0,3	0,0	8,54E-01	9	0,3	0,0	1,00E+00	6	0,2	-0,1	7,07E-01	36	0,5	0,2	5,32E-02
5-fermentation	46	0,1	5	0,2	0,1	2,44E-01	0	0,0	-0,1	7,78E-02	4	0,2	0,0	5,45E-01	4	0,1	-0,1	9,25E-02
6-gluconeogenesis/ glyoxylate cycle	25	0,1	3	0,1	0,1	2,48E-01	3	0,1	0,0	4,53E-01	1	0,0	0,0	1,00E+00	7	0,1	0,0	4,79E-01
7-oxidative pentose phosphate cycle	42	0,1	3	0,1	0,0	7,56E-01	7	0,2	0,1	8,70E-02	8	0,3	0,2	9,99E-03	3	0,0	-0,1	7,27E-02
8-TCA / org. transformation	104	0,3	2	0,1	-0,2	6,77E-02	8	0,3	0,0	1,00E+00	9	0,4	0,1	4,25E-01	42	0,5	0,3	3,86E-04
9-mitochondrial electron transport / ATP synthesis	216	0,6	13	0,5	0,0	8,90E-01	12	0,4	-0,2	3,04E-01	6	0,2	-0,3	3,33E-02	91	1,2	0,6	4,53E-08
10-cell wall	1038	2,8	63	2,6	-0,2	5,68E-01	73	2,5	-0,3	4,11E-01	98	4,1	1,3	5,05E-04	156	2,0	-0,8	1,52E-04
11-lipid metabolism	752	2,0	34	1,4	-0,6	2,96E-02	99	3,4	1,4	3,52E-06	70	2,9	0,9	4,88E-03	145	1,9	-0,1	5,01E-01
12-N-metabolism	51	0,1	5	0,2	0,1	3,95E-01	6	0,2	0,1	3,05E-01	8	0,3	0,2	2,56E-02	10	0,1	0,0	1,00E+00
13-amino acid metabolism	457	1,2	22	0,9	-0,3	1,80E-01	71	2,4	1,2	4,82E-07	31	1,3	0,1	7,75E-01	90	1,2	-0,1	7,32E-01
14-S-assimilation	18	0,0	0	0,0	0,0	6,24E-01	0	0,0	0,0	6,37E-01	3	0,1	0,1	1,32E-01	4	0,1	0,0	7,81E-01
15-metal handling	136	0,4	4	0,2	-0,2	1,13E-01	10	0,3	0,0	1,00E+00	6	0,2	-0,1	4,79E-01	28	0,4	0,0	1,00E+00
16-secondary metabolism	1028	2,8	42	1,7	-1,0	1,25E-03	96	3,3	0,5	1,02E-01	110	4,6	1,8	1,50E-06	146	1,9	-0,9	1,04E-05
17-hormone metabolism	1274	3,4	54	2,2	-1,2	7,50E-04	91	3,1	-0,3	4,26E-01	118	4,9	1,5	2,49E-04	198	2,6	-0,8	1,39E-04
18-Co-factor and vitamine metabolism	78	0,2	4	0,2	0,0	8,19E-01	5	0,2	0,0	8,33E-01	6	0,2	0,0	6,44E-01	21	0,3	0,1	2,84E-01
19-tetrapyrrole synthesis	69	0,2	2	0,1	-0,1	3,26E-01	12	0,4	0,2	1,59E-02	2	0,1	-0,1	3,26E-01	10	0,1	-0,1	3,69E-01
20-stress	3014	8,1	153	6,2	-1,9	8,08E-04	133	4,6	-3,5	3,80E-13	215	8,9	0,8	1,44E-01	526	6,9	-1,2	2,56E-04
21-redox.regulation	321	0,9	16	0,7	-0,2	3,07E-01	29	1,0	0,1	4,68E-01	32	1,3	0,5	2,44E-02	68	0,9	0,0	8,39E-01
22-polyamine metabolism	20	0,1	4	0,2	0,1	5,74E-02	2	0,1	0,0	6,72E-01	0	0,0	-0,1	6,31E-01	3	0,0	0,0	7,85E-01
23-nucleotide metabolism	260	0,7	14	0,6	-0,1	5,30E-01	38	1,3	0,6	6,89E-04	15	0,6	-0,1	8,00E-01	58	0,8	0,1	5,50E-01
24-Biodegradation of Xenobiotics	48	0,1	2	0,1	0,0	7,69E-01	5	0,2	0,0	4,35E-01	7	0,3	0,2	4,79E-02	8	0,1	0,0	7,23E-01
25-C1-metabolism	49	0,1	2	0,1	0,0	7,70E-01	10	0,3	0,2	9,45E-03	2	0,1	0,0	7,69E-01	18	0,2	0,1	4,88E-02
26-misc	2886	7,7	176	7,2	-0,6	3,29E-01	208	7,1	-0,6	2,48E-01	282	11,7	4,0	4,70E-11	458	6,0	-1,8	3,79E-08
27-RNA	6362	17,1	543	22,1	5,1	4,78E-10	560	19,2	2,1	3,68E-03	319	13,2	-3,8	6,42E-07	1552	20,3	3,2	6,17E-11
28-DNA	2498	6,7	229	9,3	2,6	1,66E-06	146	5,0	-1,7	2,63E-04	44	1,8	-4,9	7,91E-28	581	7,6	0,9	6,32E-03
29-protein	6932	18,6	540	22,0	3,4	3,96E-05	602	20,7	2,0	7,23E-03	350	14,5	-4,1	3,20E-07	1625	21,2	2,6	1,91E-07
30-signalling	4008	10,8	206	8,4	-2,4	1,70E-04	195	6,7	-4,1	3,30E-13	223	9,3	-1,5	2,05E-02	643	8,4	-2,4	2,51E-10
31-cell	1391	3,7	61	2,5	-1,2	1,02E-03	112	3,8	0,1	7,61E-01	94	3,9	0,2	6,58E-01	356	4,6	0,9	2,43E-04
32-micro RNA, natural antisense etc	2	0,0	0	0,0	0,0	1,00E+00	1	0,0	0,0	2,02E-01	0	0,0	0,0	1,00E+00	1	0,0	0,0	4,30E-01
33-development	881	2,4	54	2,2	-0,2	6,80E-01	89	3,1	0,7	2,39E-02	69	2,9	0,5	1,30E-01	150	2,0	-0,4	2,94E-02
34-transport	2088	5,6	146	6,0	0,3	4,69E-01	147	5,0	-0,6	2,24E-01	216	9,0	3,4	1,41E-10	447	5,8	0,2	4,31E-01
TOTAL	37248	100	2452	100			2914	100			2408	100			7663	100		

Annexe III-7 – Répartition des sondes au sein des classes fonctionnelles BIN dans l’analyse relative aux gènes exprimés dans les différents tissus de la graine. Les différences de représentativité des classes fonctionnelles ont été calculées afin de déterminer les classes surreprésentées (enrichies en gènes) en rouge et sous-représentées (appauvries en gènes) en bleu (cf figure III-7). Les valeurs p de probabilité (test de Fisher exact) associées à ces différences sont indiquées et un code couleur est attribué en fonction des valeurs. La classe "35.not assigned" n’a pas été prise en compte pour les calculs et ne figurent pas dans ce tableau.

CHAPITRE -3- Annexes

BinCode-BinName	ensemble des sondes la puce NimbleGen Medtr_v1.0.		analyse des sondes spécifiquement et préférentiellement exprimées dans l'albume et l'embryon		estimation de la dérégulation d'une classe (enrichissement ou appauvrissement en gènes)	valeur p de probabilité associée à la dérégulation de la classe	analyse des sondes spécifiquement et préférentiellement exprimées dans l'albume et le tégument		estimation de la dérégulation d'une classe (enrichissement ou appauvrissement en gènes)	valeur p de probabilité associée à la dérégulation de la classe	analyse des sondes spécifiquement et préférentiellement exprimées dans l'embryon et le tégument		estimation de la dérégulation d'une classe (enrichissement ou appauvrissement en gènes)	valeur p de probabilité associée à la dérégulation de la classe
	nombre de sondes	% répartition	nombre de sondes	% répartition			nombre de sondes	% répartition			nombre de sondes	% répartition		
1-photosynthesis	548	1,5	14	1,1	-0,4	2,55E-04	9	0,4	-1,1	3,95E-07	228	7,2	5,7	2,44E-71
2-major CHO metabolism	241	0,6	8	0,6	0,0	7,79E-02	13	0,5	-0,1	6,91E-01	27	0,8	0,2	1,73E-01
3-minor CHO metabolism	246	0,7	7	0,5	-0,1	4,15E-02	12	0,5	-0,2	4,31E-01	35	1,1	0,4	7,23E-03
4-glycolysis	119	0,3	5	0,4	0,1	5,62E-01	7	0,3	0,0	1,00E+00	16	0,5	0,2	1,06E-01
5-fermentation	46	0,1	4	0,3	0,2	5,36E-01	5	0,2	0,1	2,36E-01	8	0,3	0,1	7,15E-02
6-gluconeogenese/ glyoxylate cycle	25	0,1	0	0,0	-0,1	3,99E-01	5	0,2	0,1	3,25E-02	1	0,0	0,0	7,18E-01
7-oxidative pentose phosphate cycle	42	0,1	1	0,1	0,0	5,17E-01	3	0,1	0,0	7,52E-01	11	0,3	0,2	2,46E-03
8-TCA/ org. transformation	104	0,3	2	0,2	-0,1	9,44E-02	9	0,4	0,1	4,24E-01	7	0,2	-0,1	7,23E-01
9-mitochondrial electron transport / ATP synthesis	216	0,6	9	0,7	0,1	3,15E-01	10	0,4	-0,2	3,99E-01	8	0,3	-0,3	1,25E-02
10-cell wall	1038	2,8	29	2,2	-0,5	2,14E-06	86	3,6	0,8	2,61E-02	91	2,9	0,1	8,23E-01
11-lipid metabolism	752	2,0	23	1,8	-0,2	2,55E-04	72	3,0	1,0	1,86E-03	86	2,7	0,7	1,15E-02
12-N-metabolism	51	0,1	0	0,0	-0,1	7,44E-02	3	0,1	0,0	1,00E+00	6	0,2	0,1	4,55E-01
13-amino acid metabolism	457	1,2	39	3,0	1,8	5,35E-02	22	0,9	-0,3	2,09E-01	58	1,8	0,6	6,37E-03
14-S-assimilation	18	0,0	0	0,0	0,0	6,22E-01	0	0,0	0,0	6,23E-01	6	0,2	0,1	9,28E-03
15-metal handling	136	0,4	4	0,3	-0,1	1,50E-01	12	0,5	0,1	2,97E-01	10	0,3	-0,1	7,59E-01
16-secondary metabolism	1028	2,8	29	2,2	-0,5	2,72E-06	66	2,8	0,0	1,00E+00	83	2,6	-0,2	6,52E-01
17-hormone metabolism	1274	3,4	26	2,0	-1,4	1,40E-11	113	4,7	1,3	1,30E-03	118	3,7	0,3	3,89E-01
18-Co-factor and vitamine metabolism	78	0,2	6	0,5	0,3	6,36E-01	8	0,3	0,1	2,50E-01	14	0,4	0,2	1,74E-02
19-tetrapyrrole synthesis	69	0,2	3	0,2	0,0	8,00E-01	2	0,1	-0,1	4,47E-01	21	0,7	0,5	5,18E-06
20-stress	3014	8,1	55	4,3	-3,8	6,81E-30	161	6,7	-1,4	1,61E-02	215	6,8	-1,3	7,11E-03
21-redox.regulation	321	0,9	6	0,5	-0,4	7,75E-04	25	1,0	0,2	3,63E-01	53	1,7	0,8	2,92E-05
22-polyamine metabolism	20	0,1	1	0,1	0,0	1,00E+00	6	0,3	0,2	3,93E-03	0	0,0	-0,1	3,99E-01
23-nucleotide metabolism	260	0,7	21	1,6	0,9	2,47E-01	23	1,0	0,3	1,66E-01	27	0,8	0,1	3,22E-01
24-Biodegradation of Xenobiotics	48	0,1	1	0,1	-0,1	3,68E-01	2	0,1	0,0	7,69E-01	11	0,3	0,2	5,84E-03
25-C1-metabolism	49	0,1	0	0,0	-0,1	1,17E-01	1	0,0	-0,1	3,70E-01	9	0,3	0,2	4,52E-02
26-misc	2886	7,7	66	5,1	-2,6	4,27E-22	217	9,0	1,3	2,53E-02	300	9,4	1,7	9,98E-04
27-RNA	6362	17,1	281	21,7	4,7	3,49E-10	381	15,9	-1,2	1,37E-01	466	14,6	-2,4	3,51E-04
28-DNA	2498	6,7	111	8,6	1,9	2,73E-04	62	2,6	-4,1	7,32E-19	60	1,9	-4,8	7,51E-35
29-protein	6932	18,6	261	20,2	1,6	2,73E-20	516	21,5	2,9	5,53E-04	513	16,1	-2,5	4,22E-04
30-signalling	4008	10,8	72	5,6	-5,2	1,36E-40	140	5,8	-4,9	3,23E-16	295	9,3	-1,5	8,41E-03
31-cell	1391	3,7	84	6,5	2,8	9,10E-01	116	4,8	1,1	8,10E-03	145	4,6	0,8	2,31E-02
32-micro RNA, natural antisense etc	2	0,0	0	0,0	0,0	1,00E+00	0	0,0	0,0	1,00E+00	0	0,0	0,0	1,00E+00
33-development	881	2,4	44	3,4	1,0	1,77E-01	99	4,1	1,8	6,14E-07	86	2,7	0,3	2,27E-01
34-transport	2088	5,6	81	6,3	0,7	8,64E-06	193	8,0	2,4	2,37E-06	171	5,4	-0,2	6,01E-01
TOTAL	37248	100	1293	100			2399	100			3185	100		

 p < 0,001
 p < 0,01
 p < 0,05

Annexe III-7 – suite.

Conclusion générale et perspectives

Le développement et la structure de l'albumen jouent un rôle crucial dans la formation de la graine. L'objectif de la thèse était de caractériser plus précisément l'importance de ce tissu dans cette étape clé du cycle de vie des plantes qu'est le développement de la graine. Deux approches ont été utilisées : une ciblée sur un FT de type DOF préférentiellement exprimé dans les graines au niveau de l'albumen chalazal (chapitre 2) et une globale d'analyse du transcriptome des trois tissus de la graine au stade clé de transition entre l'embryogenèse (importantes divisions cellulaires) et le remplissage (expansion des cellules de l'embryon et accumulation des réserves) (chapitre 3).

Des mutants TILLING et Tnt1 pour un FT de type DOF ont été isolés et caractérisés, montrant un effet létal de la mutation à l'état homozygote sur le développement de la graine. Des plantes porteuses d'une mutation EMS à l'état homozygote (*dash* EMS109) ont toutefois été obtenues. Celles-ci présentent un phénotype d'avortement de la totalité des gousses durant les premières semaines. Après 10 à 12 semaines, les gousses se développent avant d'avorter vers 8-10 jap. Enfin, un processus d'échappement à l'avortement se met en place en fin de cycle de vie afin de permettre à ces plantes mutantes de produire des gousses et graines, plus petites que les gousses et graines des plantes de génotype sauvage. Afin de caractériser plus précisément les mécanismes moléculaires à l'origine de l'avortement, le transcriptome à 8 jap (et 10 jap) des gousses du mutant *dash* EMS109 a été comparé à celui des gousses du génotype sauvage (chapitre 2). Cette étude a révélé de nombreuses classes fonctionnelles dérégulées chez le mutant, notamment la classe relative au métabolisme hormonal (tableau II-2). Plusieurs pistes ont ensuite été suivies : étude globale des familles de gènes dérégulés à l'intérieur de cette classe (cf. article en préparation, chapitre 2) puis étude plus ciblée des gènes intervenant dans les voies de biosynthèse (cf. paragraphe II.3.3, chapitre 2). Ces approches ont été complétées par des tentatives de complémentation par application exogène d'auxine (AIA et TIBA paragraphe II.2.2.9 et 2-4D paragraphe II.3.4) et par des dosages d'auxines et auxines conjuguées dans les gousses (Coll. J Verdier, The Samuel Roberts Noble Foundation, paragraphe II.2.2.9). Ceci nous a permis de constater une forte accumulation d'auxine et de ses formes conjuguées AIA-Asp dans les gousses, suggérant une non-utilisation des hormones sous forme libre (AIA), les formes conjuguées AIA-Asp visant à être dégradées (les voies de dégradation liées à l'ubiquitine étant d'ailleurs fortement sur-exprimés dans les gousses du mutant *dash* EMS109, paragraphe II.2.2.7). De plus, des applications exogènes d'auxine ont conduit à une restauration partielle du phénotype, les gousses du mutant *dash* EMS109 obtenues en début de cycle se développent et produisent des graines de poids comparables aux gousses se développant en fin de cycle. Le mécanisme d'échappement à l'avortement intervient donc plus tôt lorsque l'on applique de l'auxine exogène à des gousses vouées à l'avortement. De nombreuses interrogations sont donc soulevées suite à ces résultats et notamment comment un apport d'auxine dans une gousse qui en contient déjà trop parvient à restaurer le phénotype ? Plusieurs hypothèses sont donc formulées en proposant qu'il puisse y avoir des troubles dans le transport de l'auxine entre la graine et la gousse, conduisant à un avortement précoce des gousses, ou

au sein même de la graine, empêchant le développement de l'embryon et conduisant à des défauts dans la structure de l'albumen (figure II-5 et annexe II-3 et II-4, paragraphe II.2.2.4). On pourrait également penser qu'il s'agit de troubles dans la perception de l'auxine au niveau de l'embryon, les gènes de réponse à l'auxine (de type SAUR ou Aux/AIA) étant en majorité sous-exprimés dans les gousses du mutant *dash* EMS109. Une baisse d'expression dans les graines isolées du mutant *dash* EMS109 a été confirmée pour un gène sur trois (tableau II-4, paragraphe II.3.2). Il serait intéressant d'élargir cette étude au transcriptome de gousses isolées chez les mutants *dash* EMS109 par rapport aux gousses isolées provenant de génotype sauvage pour répondre à certaines interrogations quant aux communications hormonales entre la gousse et la graine. Ceci permettrait de déterminer si la perception de l'auxine est affectée dans la gousse. L'étude de la dérégulation de ces gènes chez le mutant *dash* NF6042 serait également envisageable afin de valider l'importance des dérégulations dans le développement des graines. Les gènes cibles ainsi déterminés pourraient ensuite faire l'objet d'une validation fonctionnelle par des techniques telle que le double hybride ou le gel retard. Cette dernière approche (*i.e.* gel retard) a été envisagée en vue de mettre en évidence les interactions possibles de DASH avec les gènes cibles potentiels identifiés suite aux analyses transcriptomiques. Pour cela, le clonage du gène *DASH* dans un vecteur de clonage (pET41a) permettant l'expression de la protéine DASH fusionnée à un tag histidine a été réalisé. L'expression de la protéine DASH a été vérifiée mais les cibles potentielles ayant été identifiées tardivement (comme par exemple les défensines, figure II-6, paragraphe II.2.2.6), aucune expérience de validation par gel retard n'a pu être entreprise dans le cadre de la thèse. L'étude présentée dans le chapitre 3 a permis d'étudier la spécificité tissulaire des cibles potentielles de DASH. Il est à noter que cette étude a été réalisée à un seul stade de développement (12 jap) pour des questions de faisabilité technique (quantités de matériel biologique et difficulté d'isoler l'albumen syncytial, trop aqueux, dans des graines de *M. truncatula* à 8 ou 10 jap). A ce stade, *DASH* est très fortement exprimé (figure II-1), nous permettant ainsi de mettre plus facilement en évidence les localisations tissulaires de ses gènes cibles putatifs (figure II-6 et tableau II-4). Ainsi on observe une grande part de gènes dérégulés ayant une expression préférentielle dans l'albumen ou dans le tégument, ce qui est rassurant étant donné l'expression de DASH, dans l'albumen chalazal, à la jonction entre l'albumen et le tégument (figure II-1B). Ce FT pourrait donc réguler l'expression de gènes impliqués dans la communication entre ces deux tissus.

L'étude du transcriptome des trois tissus a permis de souligner l'importance de certaines classes fonctionnelles dans un tissu donné de la graine. En effet, les classes "27.RNA" et "29.protein" se distinguent par un nombre de transcrits relativement important dans l'albumen et l'embryon, suggérant une forte activité transcriptionnelle et de synthèse protéique dans ces tissus. L'analyse détaillée de ces classes a révélé une part considérable de gènes liés à la dégradation *via* l'ubiquitine dans l'albumen. Ces gènes sont susceptibles de contrôler la dégradation de l'albumen de manière progressive et coordonnée avec le remplissage de l'embryon, ou encore de transcrits codant des protéines ribosomales dans l'embryon qui, à

12 jap, se prépare à accumuler ses protéines de réserve (figure III-9). On note également une abondance du nombre de gènes de la classe "26.miscellaneous enzyme family" dans le tégument, principalement des cytochromes P450 et des UDP glucosyl et glucoronyl transférases (figure III-11). Cette étude du transcriptome des trois tissus a donc permis de pointer des spécificités tissulaires. Par ailleurs, elle représente un outil précieux pour une approche plus ciblée consistant à rechercher la localisation tissulaire d'un gène donné. Dans une perspective d'étude du développement conjoint des trois tissus, il conviendrait d'élargir cette étude du transcriptome à d'autres stades de développement pendant le remplissage/la maturation. Ceci permettrait de mettre en évidence les spécificités tissulaires à une autre étape du développement de la graine et de mettre l'accent sur des voies métaboliques préférentielles de l'embryogénèse et du remplissage ou de la maturation dans un tissu donné.

Une orientation plus générale de ce travail serait d'identifier des allèles du gène *DASH* favorables pour la taille de la graine permettant une utilisation en sélection variétale. Une perspective sera d'exploiter la variabilité génétique naturelle et induite chez le pois pour étudier les variations alléliques du gène *DASH* dans les différentes populations afin de déterminer si des allèles de ce gène sont corrélés à des variations de poids d'une graine. Le criblage TILLING de la population EMS chez le pois a déjà permis de détecter des mutations dans la séquence de *DASH* et les rétro-croisements de ces mutants avec le génotype sauvage (Caméor) sont en cours afin de purifier le fond génétique. Il sera aussi intéressant d'approfondir l'étude de ce facteur de transcription chez le pois. Ainsi, il sera plus facile de travailler sur des graines et gousses de pois, les différents tissus étant plus faciles à isoler. Des dosages d'auxines, de protéines totales ou de teneur en sucre ou en lipides seront également plus aisés à entreprendre.

Références bibliographiques

-A-

Abirached-Darmency M, Abdel-gawwad MR, Conejero G, Verdeil JL, Thompson R (2005) In situ expression of two storage protein genes in relation to histo-differentiation at mid-embryogenesis in *Medicago truncatula* and *Pisum sativum* seeds. *J Exp Bot* **56**: 2019-2028

Adamski NM, Anastasiou E, Eriksson S, O'Neill CM, Lenhard M (2009) Local maternal control of seed size by KLUH/CYP78A5-dependent growth signaling. *Proc Natl Acad Sci U S A* **106**: 20115-20120

Agarwal P, Kapoor S, Tyagi AK (2011) Transcription factors regulating the progression of monocot and dicot seed development. *Bioessays* **33**: 189-202

Ahmad M, Rim Y, Chen H, Kim JY (2013) Functional characterization of Arabidopsis Dof transcription factor AtDof4.1. *Russian Journal of Plant Physiology* **60**: 116-123

Aida M, Ishida T, Fukaki H, Fujisawa H, Tasaka M (1997) Genes involved in organ separation in Arabidopsis: an analysis of the cup-shaped cotyledon mutant. *The Plant Cell Online* **9**: 841-857

Arce AL, Cabello JV, Chan RL (2008) Patents on plant transcription factors. *Recent patents on biotechnology* **2**: 209-217

Autran D, Baroux C, Raissig MT, Lenormand T, Wittig M, Grob S, Steimer A, Barann M, Klostermeier UC, Leblanc O, Vielle-Calzada JP, Rosenstiel P, Grimanelli D, Grossniklaus U (2011) Maternal epigenetic pathways control parental contributions to Arabidopsis early embryogenesis. *Cell* **145**: 707-719

-B-

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* **37**: W202-208

Barker DG, Bianchi S, Blondon F, Dattee Y, Duc G, Essad S, Flament P, Gallusci P, Genier G, Guy P, Muel X, Tournier J, Denarie J, Huguet T (1990) *Medicago truncatula*, a model plant for studying the molecular genetics of the Rhizobium-legume symbiosis. *Plant Molecular Biology Reporter* **8**: 40-49

Batchelor AK, Boutilier K, Miller SS, Labbe H, Bowman L, Hu M, Johnson DA, Gijzen M, Miki BL (2000) The seed coat-specific expression of a subtilisin-like gene, SCS1, from soybean. *Planta* **211**: 484-492

Baumann K, Paolis Ad, Costantino P, Gualberti G (1999) The DNA binding site of the Dof protein NtBBF1 is essential for tissue-specific and auxin-regulated expression of the rolB oncogene in plants. *Plant Cell* **11**: 323-333

Bäumlein H, Miséra S, Luerßen H, Kölle K, Horstmann C, Wobus U, Müller AJ (1994) The FUS3 gene of *Arabidopsis thaliana* is a regulator of gene expression during late embryogenesis. *The Plant Journal* **6**: 379-387

Belmonte MF, Kirkbride RC, Stone SL, Pelletier JM, Bui AQ, Yeung EC, Hashimoto M, Fei J, Harada M, Munoz MD, Le BH, Drews GN, Brady SM, Goldberg RB, Harada JJ (2013) Comprehensive developmental profiles of gene activity in regions and subregions of the Arabidopsis seed. *Proceedings of the National Academy of Sciences of the United States of America* **110**: E435-E444

Benedito VA, Torres-Jerez I, Murray JD, Andrianakaja A, Allen S, Kakar K, Wandrey M, Verdier J, Zuber H, Ott T, Moreau S, Niebel A, Frickey T, Weiller G, He J, Dai X, Zhao PX, Tang Y, Udvardi MK (2008) A gene expression atlas of the model legume *Medicago truncatula*. *Plant Journal* **55**: 504-513

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**: 289-300

Berger F, Grini PE, Schnittger A (2006) Endosperm: an integrator of seed growth and development. *Current Opinion in Plant Biology* **9**: 664-670

Berry T, Bewley JD (1991) Seeds of tomato (*Lycopersicon esculentum* Mill.) which develop in a fully hydrated environment in the fruit switch from a developmental to a germinative mode without a requirement for desiccation. *Planta* **186**: 27-34

Bradley D, Carpenter R, Sommer H, Hartley N, Coen E (1993) Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the *plena* locus of *Antirrhinum*. *Cell* **72**: 85-95

Brivanlou AH, Darnell JE, Jr. (2002) Signal transduction and the control of gene expression. *Science* **295**: 813-818

Brosche M, Strid A (1999) Cloning, expression, and molecular characterization of a small pea gene family regulated by low levels of ultraviolet B radiation and other stresses. *Plant Physiol* **121**: 479-487

Buitink J, Leger JJ, Guisle I, Vu BL, Wuilleme S, Lamirault G, Le Bars A, Le Meur N, Becker A, Kuester H, Leprince O (2006) Transcriptome profiling uncovers metabolic and regulatory processes occurring during the transition from desiccation-sensitive to desiccation-tolerant stages in *Medicago truncatula* seeds. *Plant Journal* **47**: 735-750

-C-

Callis J, Vierstra RD (2000) Protein degradation in signaling. *Current Opinion in Plant Biology* **3**: 381-386

Capron D, Mouzeyar S, Boulaflous A, Girousse C, Rustenholz C, Laugier C, Paux E, Bouzidi MF (2012) Transcriptional profile analysis of E3 ligase and hormone-related genes expressed during wheat grain development. *Bmc Plant Biology* **12**

Cavaillès E (2009) La relance des légumineuses dans le cadre d'un plan protéine: quels bénéfices environnementaux? *Etudes & documents - Commissariat Général du Développement Durable* n°15

Cavalar M, Phlippen Y, Kreuzaler F, Peterhaensel C (2007) A drastic reduction in DOF1 transcript levels does not affect C-4-specific gene expression in maize. *Journal of Plant Physiology* **164**: 1665-1674

Chatelain E, Hundertmark M, Leprince O, Le Gall S, Satour P, Deligny-Penninck S, Rogniaux H, Buitink J (2012) Temporal profiling of the heat-stable proteome during late maturation of *Medicago truncatula* seeds identifies a restricted subset of late embryogenesis abundant proteins associated with longevity. *Plant Cell Environ* **35**: 1440-1455

Chaudhury AM, Ming L, Miller C, Craig S, Dennis ES, Peacock WJ (1997) Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* **94**: 4223-4228

Chen WQ, Chao G, Singh KB (1996) The promoter of a H₂O₂-inducible, *Arabidopsis* glutathione S-transferase gene contains closely linked OBF- and OBP1-binding sites. *Plant Journal* **10**: 955-966

Chen X, Wang D, Liu C, Wang M, Wang T, Zhao Q, Yu J (2012) Maize transcription factor Zmdof1 involves in the regulation of Zm401 gene. *Plant Growth Regulation* **66**: 271-284

Chen Y, Xu Y, Luo W, Li W, Chen N, Zhang D, Chong K (2013) The F-Box Protein OsFBK12 Targets OsSAMS1 for Degradation and Affects Pleiotropic Phenotypes, Including Leaf Senescence, in Rice. *Plant Physiology* **163**: 1673-1685

Cheng Y, Dai X, Zhao Y (2007) Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in *Arabidopsis*. *Plant Cell* **19**: 2430-2439

Choi HK, Mun JH, Kim DJ, Zhu HY, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004) Estimating genome conservation between crop and model legume species. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 15289-15294

Choi YE, Kim HS, Soh WY, Yang DC (1997) Developmental and structural aspects of somatic embryos formed on medium containing 2,3,5-triiodobenzoic acid. *Plant Cell Reports* **16**: 738-744

Churchill GA (2002) Fundamentals of experimental design for cDNA microarrays. *Nature Genetics* **32**: 490-495

Coe EH, Jr., Neuffer MG, Hoisington DA (1988) The genetics of corn. *In* Corn and corn improvement. Third edition., pp 81-258

Coen ES, Romero JM, Doyle S, Elliott R, Murphy G, Carpenter R (1990) *floricaula*: a homeotic gene required for flower development in *antirrhinum majus*. *Cell* **63**: 1311-1322

Cominelli E, Galbiati M, Albertini A, Fornara F, Conti L, Coupland G, Tonelli C (2011) DOF-binding sites additively contribute to guard cell-specificity of AtMYB60 promoter. *Bmc Plant Biology* **11**

Cossegal M, Chambrier P, Mbalo S, Balzergue S, Martin-Magniette ML, Moing A, Deborde C, Guyon V, Perez P, Rogowsky P (2008) Transcriptional and metabolic adjustments in ADP-glucose pyrophosphorylase-deficient bt2 maize kernels. *Plant Physiol* **146**: 1553-1570

Cosson V, Durand P, d'Erfurth I, Kondorosi A, Ratet P (2006) *Medicago truncatula* transformation using leaf explants. *Methods Mol Biol* **343**: 115-127

Curtis MD, Grossniklaus U (2003) A gateway cloning vector set for high-throughput functional analysis of genes in planta. *Plant Physiology* **133**: 462-469

-D-

d'Erfurth I, Cosson V, Eschstruth A, Lucas H, Kondorosi A, Ratet P (2003) Efficient transposition of the Tnt1 tobacco retrotransposon in the model legume *Medicago truncatula*. *Plant Journal* **34**: 95-106

D'Erfurth I, Le Signor C, Aubert G, Sanchez M, Vernoud V, Darchy B, Lherminier J, Bourion V, Bouteiller N, Bendahmane A, Buitink J, Prosperi JM, Thompson R, Burstin J, Gallardo K (2012) A role for an endosperm-localized subtilase in the control of seed size in legumes. *New Phytologist* **196**: 738-751

Day RC, Herridge RP, Ambrose BA, Macknight RC (2008) Transcriptome Analysis of Proliferating *Arabidopsis* Endosperm Reveals Biological Implications for the Control of Syncytial Division, Cytokinin Signaling, and Gene Expression Regulation. *Plant Physiology* **148**: 1964-1984

de Castro E, Sigrist CJA, Gattiker A, Bulliard V, Langendijk-Genevaux PS, Gasteiger E, Bairoch A, Hulo N (2006) ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research* **34**: W362-W365

Delbarre A, Muller P, Imhoff V, Guern J (1996) Comparison of mechanisms controlling uptake and accumulation of 2,4-dichlorophenoxy acetic acid, naphthalene-1-acetic acid, and indole-3-acetic acid in suspension-cultured tobacco cells. *Planta* **198**: 532-541

DePaolis A, Sabatini S, DePascalis L, Costantino P, Capone I (1996) A rolB regulatory factor belongs to a new class of single zinc finger plant proteins. *Plant Journal* **10**: 215-223

Despres C, Chubak C, Rochon A, Clark R, Bethune T, Desveaux D, Fobert PR (2003) The *Arabidopsis* NPR1 disease resistance protein is a novel cofactor that confers redox regulation of DNA binding activity to the basic domain/leucine zipper transcription factor TGA1. *Plant Cell* **15**: 2181-2191

Dharmasiri N, Dharmasiri S, Estelle M (2005) The F-box protein TIR1 is an auxin receptor. *Nature* **435**: 441-445

Diaz I, Vicente-Carbajosa J, Abraham Z, Martinez M, Isabel-La Moneda I, Carbonero P (2002) The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. *Plant Journal* **29**: 453-464

Djemel N, Guedon D, Lechevalier A, Salon C, Miquel M, Prosperi JM, Rochat C, Boutin JP (2005) Development and composition of the seeds of nine genotypes of the *Medicago truncatula* species complex. *Plant Physiol Biochem* **43**: 557-566

Dong G, Ni Z, Yao Y, Nie X, Sun Q (2007) Wheat Dof transcription factor WPBF interacts with TaQM and activates transcription of an alpha-gliadin gene during wheat seed development. *Plant Molecular Biology* **63**: 73-84

Dumas C, Rogowsky P (2008) Fertilization and early seed formation. *Comptes Rendus Biologies* **331**: 715-725

-E-

Engel ML, Davis RH, McCormick S (2005) Green sperm. Identification of male gamete promoters in Arabidopsis. *Plant Physiology* **138**: 2124-2133

Esaka M, Fujisawa K, Goto M, Kisu Y (1992) Regulation of Ascorbate Oxidase Expression in Pumpkin by Auxin and Copper. *Plant Physiology* **100**: 231-237

-F-

Fang W, Wang Z, Cui R, Li J, Li Y (2012) Maternal control of seed size by EOD3/CYP78A6 in Arabidopsis thaliana. *Plant J* **70**: 929-939

Feller A, Machemer K, Braun EL, Grotewold E (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant Journal* **66**: 94-116

Finnegan EJ, Peacock WJ, Dennis ES (2000) DNA methylation, a key regulator of plant development and other processes. *Curr Opin Genet Dev* **10**: 217-223

Fiume E, Fletcher JC (2012) Regulation of Arabidopsis embryo and endosperm development by the polypeptide signaling molecule CLE8. *Plant Cell* **24**: 1000-1012

Forestan C, Meda S, Varotto S (2010) ZmPIN1-mediated auxin transport is related to cellular differentiation during maize embryogenesis and endosperm development. *Plant Physiol* **152**: 1373-1390

Forestan C, Varotto S (2011) The role of PIN auxin efflux carriers in polar auxin transport and accumulation and their effect on shaping maize development. *Mol Plant* **5**: 787-798

Friml J, Vieten A, Sauer M, Weijers D, Schwarz H, Hamann T, Offringa R, Jurgens G (2003) Efflux-dependent auxin gradients establish the apical-basal axis of Arabidopsis. *Nature* **426**: 147-153

Frugoli J, Harris J (2001) Medicago truncatula on the move! *The Plant cell* **13**: 458-463

-G-

- Gabriele S, Rizza A, Martone J, Circelli P, Costantino P, Vittorioso P (2010) The Dof protein DAG1 mediates PIL5 activity on seed germination by negatively regulating GA biosynthetic gene AtGA3ox1. *Plant Journal* **61**: 312-323
- Gagnot S, Tamby JP, Martin-Magniette ML, Bitton F, Taconnat L, Balzergue S, Aubourg S, Renou JP, Lecharny A, Brunaud V (2008) CATdb: a public access to Arabidopsis transcriptome data from the URGV-CATMA platform. *Nucleic Acids Res* **36**: D986-990
- Gallardo K, Firnhaber C, Zuber H, Hericher D, Belghazi M, Henry C, Kuster H, Thompson R (2007) A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues. *Mol Cell Proteomics* **6**: 2165-2179
- Gallardo K, Le Signor C, Vandekerckhove J, Thompson RD, Burstin J (2003) Proteomics of *Medicago truncatula* seed development establishes the time frame of diverse metabolic processes related to reserve accumulation. *Plant Physiol* **133**: 664-682
- Gallardo K, Thompson R, Burstin J (2008) Reserve accumulation in legume seeds. *C R Biol* **331**: 755-762
- Garcia D, Fitz Gerald JN, Berger F (2005) Maternal control of integument cell elongation and zygotic control of endosperm growth are coordinated to determine seed size in *Arabidopsis*. *Plant Cell* **17**: 52-60
- Garcia D, Saingery V, Chambrier P, Mayer U, Jurgens G, Berger F (2003) *Arabidopsis* haiku mutants reveal new controls of seed size by endosperm. *Plant Physiol* **131**: 1661-1670
- Gaur VS, Singh US, Kumar A (2011) Transcriptional profiling and in silico analysis of Dof transcription factor gene family for understanding their regulation during seed development of rice *Oryza sativa* L. *Molecular Biology Reports* **38**: 2827-2848
- Geisler-Lee J, Gallie DR (2005) Aleurone cell identity is suppressed following connation in maize kernels. *Plant Physiol* **139**: 204-212
- Gillman JD, Tetlow A, Lee J-D, Shannon JG, Bilyeu K (2011) Loss-of-function mutations affecting a specific Glycine max R2R3 MYB transcription factor result in brown hilum and brown seed coats. *Bmc Plant Biology* **11**
- Goldberg RB, de Paiva G, Yadegari R (1994) Plant embryogenesis: zygote to seed. *Science* **266**: 605-614
- Gomez-Ros LV, Gabaldon C, Nunez-Flores MJL, Gutierrez J, Herrero J, Zapata JM, Sottomayor M, Cuello J, Barcelo AR (2012) The promoter region of the *Zinnia elegans* basic peroxidase isoenzyme gene contains cis-elements responsive to nitric oxide and hydrogen peroxide. *Planta* **236**: 327-342
- Graur D, Li W (2000) *Fundamentals of molecular evolution.*, Ed 2nd Edition. Sinauer Associates, Stamford

Grunewald W, De Smet I, De Rybel B, Robert HS, van de Cotte B, Willemsen V, Gheysen G, Weijers D, Friml J, Beeckman T (2013) Tightly controlled WRKY23 expression mediates Arabidopsis embryo development. *EMBO Rep* **14**: 1136-1142

Grunewald W, De Smet I, Lewis DR, Lofke C, Jansen L, Goeminne G, Vanden Bossche R, Karimi M, De Rybel B, Vanholme B, Teichmann T, Boerjan W, Van Montagu MCE, Gheysen G, Muday GK, Friml J, Beeckman T (2012) Transcription factor WRKY23 assists auxin distribution patterns during Arabidopsis root development through local control on flavonol biosynthesis. *Proceedings of the National Academy of Sciences* **109**: 1554-1559

Gualberti G, Papi M, Bellucci L, Ricci I, Bouchez D, Camilleri C, Costantino P, Vittorioso P (2002) Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of Arabidopsis seeds. *Plant Cell* **14**: 1253-1263

Guo Y, Qiu LJ (2013) Genome-wide analysis of the Dof transcription factor gene family reveals soybean-specific duplicable and functional characteristics. *PLoS One* **8**: e76809

Gustafson FG (1939) Auxin distribution in fruits and its significance in fruit development. *Am. J. Bot.* **26**: 189-194

-H-

Hamamoto H, Shishido Y, Furuya S, Yasuba K (1998) Growth and Development of Tomato Fruit as Affected by 2, 3, 5-Triiodobenzoic Acid (TIBA) Applied to the Peduncle. *Journal of the Japanese Society for Horticultural Science* **67**: 210-212

Haque MA, Sasaki N, Kanegae H, Mimori S, Gao JS, Nyunoya H (2008) Identification of a Tobacco mosaic virus elicitor-responsive sequence in the resistance gene N. *Physiological and Molecular Plant Pathology* **73**: 101-108

Haughn G, Chaudhury A (2005) Genetic analysis of seed coat development in Arabidopsis. *Trends in Plant Science* **10**: 472-477

He J, Benedito VA, Wang M, Murray JD, Zhao PX, Tang Y, Udvardi MK (2009) The Medicago truncatula gene expression atlas web server. *BMC Bioinformatics* **10**: 441

Hehenberger E, Kradolfer D, Kohler C (2012) Endosperm cellularization defines an important developmental transition for embryo development. *Development* **139**: 2031-2039

Hernando-Amado S, Gonzalez-Calle V, Carbonero P, Barrero-Sicilia C (2012) The family of DOF transcription factors in Brachypodium distachyon: phylogenetic comparison with rice and barley DOFs and expression profiling. *BMC Plant Biology* **12**: (5 November 2012)-(2015 November 2012)

Hill LM, Morley-Smith ER, Rawsthorne S (2003) Metabolism of sugars in the endosperm of developing seeds of oilseed rape. *Plant Physiol* **131**: 228-236

Hoecker U, Vasil IK, McCarty DR (1995) Integrated Control of Seed Maturation and Germination Programs by Activator and Repressor Functions of Viviparous-1 of Maize. *Genes & Development* **9**: 2459-2469

Hoecker U, Vasil IK, McCarty DR (1999) Signaling from the embryo conditions Vp1-mediated repression of alpha-amylase genes in the aleurone of developing maize seeds. *Plant Journal* **19**: 371-377

Holdsworth MJ, Bentsink L, Soppe WJJ (2008) Molecular networks regulating Arabidopsis seed maturation, after-ripening, dormancy and germination. *New Phytologist* **179**: 33-54

Hong JK, Choi HW, Hwang IS, Kim DS, Kim NH, Choi DS, Kim YJ, Hwang BK (2008) Function of a novel GDSL-type pepper lipase gene, CaGLIP1, in disease susceptibility and abiotic stress tolerance. *Planta* **227**: 539-558

Hou BK, Lim EK, Higgins GS, Bowles DJ (2004) N-glucosylation of cytokinins by glycosyltransferases of *Arabidopsis thaliana*. *Journal of Biological Chemistry* **279**: 47822-47832

Hu YF, Li YP, Zhang JJ, Liu HM, Tian ML, Huang YB (2012) Binding of ABI4 to a CACCG motif mediates the ABA-induced expression of the ZmSSI gene in maize (*Zea mays* L.) endosperm. *Journal of Experimental Botany* **63**: 5979-5989

Hudson M, Ringli C, Boylan MT, Quail PH (1999) The FAR1 locus encodes a novel nuclear protein specific to phytochrome A signaling. *Genes Dev* **13**: 2017-2027

Huh JH, Bauer MJ, Hsieh TF, Fischer R (2007) Endosperm gene imprinting and seed development. *Curr Opin Genet Dev* **17**: 480-485

-|-

Ingram GC (2010) Family life at close quarters: communication and constraint in angiosperm seed development. *Protoplasma* **247**: 195-214

Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**: 249-264

Isabel-LaMoneda I, Diaz I, Martinez M, Mena M, Carbonero P (2003) SAD: a new DOF protein from barley that activates transcription of a cathepsin B-like thiol protease gene in the aleurone of germinating seeds. *Plant Journal* **33**: 329-340

Iwamoto M, Higo K, Takano M (2009) Circadian clock- and phytochrome-regulated Dof-like gene, Rdd1, is associated with grain size in rice. *Plant Cell and Environment* **32**: 592-603

-J-

Jain M, Nijhawan A, Arora R, Agarwal P, Ray S, Sharma P, Kapoor S, Tyagi AK, Khurana JP (2007) F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiology* **143**: 1467-1483

Jin J, Zhang H, Kong L, Gao G, Luo J (2014) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Research* **42**: D1182-D1187

Jin J, Zhang H, Kong L, Gao G, Luo J (2014) PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res* **42**: D1182-1187

-K-

Kang HG, Foley RC, Onate-Sanchez L, Lin CGT, Singh KB (2003) Target genes for OBP3, a Dof transcription factor, include novel basic helix-loop-helix domain proteins inducible by salicylic acid. *Plant Journal* **35**: 362-372

Kang HG, Singh KB (2000) Characterization of salicylic acid-responsive, arabidopsis Dof domain proteins: overexpression of OBP3 leads to growth defects. *Plant Journal* **21**: 329-339

Kant S, Bi YM, Zhu T, Rothstein SJ (2009) SAUR39, a small auxin-up RNA gene, acts as a negative regulator of auxin synthesis and transport in rice. *Plant Physiol* **151**: 691-701

Kant S, Rothstein S (2009) Auxin-responsive SAUR39 gene modulates auxin level in rice. *Plant Signal Behav* **4**: 1174-1175

Kawakatsu T, Takaiwa F (2010) Differences in Transcriptional Regulatory Mechanisms Functioning for Free Lysine Content and Seed Storage Protein Accumulation in Rice Grain. *Plant and Cell Physiology* **51**: 1964-1974

Kepinski S, Leyser O (2005) The Arabidopsis F-box protein TIR1 is an auxin receptor. *Nature* **435**: 446-451

Kim HS, Kim SJ, Abbasi N, Bressan RA, Yun DJ, Yoo SD, Kwon SY, Choi SB (2010) The DOF transcription factor Dof5.1 influences leaf axial patterning by promoting *Revoluta* transcription in Arabidopsis. *Plant J* **64**: 524-535

Kisu Y, Esaka M, Suzuki M (1995) Putative zinc-binding domain of plant transcription factor, AOBP, is related to DNA-binding domains of steroid hormone receptors and GATA1. *Proceedings of the Japan Academy Series B-Physical and Biological Sciences* **71**: 288-292

Kisu Y, Ono T, Shimofurutani N, Suzuki M, Esaka M (1998) Characterization and expression of a new class of zinc finger protein that binds to silencer region of ascorbate oxidase gene. *Plant and Cell Physiology* **39**: 1054-1064

Kloosterman B, Abelenda JA, Gomez Mdel M, Oortwijn M, de Boer JM, Kowitzwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S, Visser RG, Bachem CW (2013) Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature* **495**: 246-250

Kondou Y, Nakazawa M, Kawashima M, Ichikawa T, Yoshizumi T, Suzuki K, Ishikawa A, Koshi T, Matsui R, Muto S, Matsui M (2008) RETARDED GROWTH OF EMBRYO1, a new basic helix-loop-helix protein, expresses in endosperm to control embryo growth. *Plant Physiology* **147**: 1924-1935

Konishi M, Yanagisawa S (2007) Sequential activation of two Dof transcription factor gene promoters during vascular development in *Arabidopsis thaliana*. *Plant Physiology and Biochemistry* **45**: 623-629

Kosugi S, Ohashi Y (1997) PCF1 and PCF2 specifically bind to cis elements in the rice proliferating cell nuclear antigen gene. *Plant Cell* **9**: 1607-1619

Kosugi S, Suzuka I, Ohashi Y (1995) Two of three promoter elements identified in a rice gene for proliferating cell nuclear antigen are essential for meristematic tissue-specific expression. *Plant J* **7**: 877-886

Krebs J, Mueller-Roeber B, Ruzicic S (2010) A novel bipartite nuclear localization signal with an atypically long linker in DOF transcription factors. *J Plant Physiol* **167**: 583-586

Krohn NM, Yanagisawa S, Grasser KD (2002) Specificity of the stimulatory interaction between chromosomal HMGB proteins and the transcription factor Dof2 and its negative regulation by protein kinase CK2-mediated phosphorylation. *J Biol Chem* **277**: 32438-32444

Kurai T, Wakayama M, Abiko T, Yanagisawa S, Aoki N, Ohsugi R (2011) Introduction of the ZmDof1 gene into rice enhances carbon and nitrogen assimilation under low-nitrogen conditions. *Plant Biotechnology Journal* **9**: 826-837

Kurdyukov S, Song Y, Sheahan M, Rose R (2014) Transcriptional regulation of early embryo development in the model legume *Medicago truncatula*. *Plant Cell Reports* **33**: 349-362

Kushwaha H, Gupta S, Singh VK, Rastogi S, Yadav D (2011) Genome wide identification of Dof transcription factor gene family in sorghum and its comparative phylogenetic analysis with rice and *Arabidopsis*. *Molecular Biology Reports* **38**: 5037-5053

-L-

Larsen K (2004) Cloning and characterization of a ryegrass (*Lolium perenne*) gene encoding cinnamoyl-CoA reductase (CCR). *Plant Science* **166**: 569-581

Lau OS, Deng XW (2010) Plant hormone signaling lightens up: integrators of light and hormones. *Curr Opin Plant Biol* **13**: 571-577

Laux T, Wurschum T, Breuninger H (2004) Genetic regulation of embryonic pattern formation. *Plant Cell* **16 Suppl**: S190-202

- Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, Drews GN, Fischer RL, Okamura JK, Harada JJ, Goldberg RB (2007) Global analysis of gene activity during Arabidopsis seed development and identification of seed-specific transcription factors. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 8063-8070
- Le Signor C, Savoie V, Aubert G, Verdier J, Nicolas M, Pagny G, Moussy F, Sanchez M, Baker D, Clarke J, Thompson R (2009) Optimizing TILLING populations for reverse genetics in *Medicago truncatula*. *Plant Biotechnol J* **7**: 430-441
- LeClere S, Schmelz EA, Chourey PS (2008) Cell wall invertase-deficient miniature1 kernels have altered phytohormone levels. *Phytochemistry* **69**: 692-699
- LeClere S, Schmelz EA, Chourey PS (2010) Sugar Levels Regulate Tryptophan-Dependent Auxin Biosynthesis in Developing Maize Kernels. *Plant Physiology* **153**: 306-318
- Li DJ, Yang CH, Li XB, Gan Q, Zhao XF, Zhu LH (2009) Functional characterization of rice OsDof12. *Planta* **229**: 1159-1169
- Li J, Li G, Gao S, Martinez C, He G, Zhou Z, Huang X, Lee JH, Zhang H, Shen Y, Wang H, Deng XW (2010) Arabidopsis transcription factor ELONGATED HYPOCOTYL5 plays a role in the feedback regulation of phytochrome A signaling. *Plant Cell* **22**: 3634-3649
- Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Xu C, Li X, Xiao J, He Y, Zhang Q (2011) Natural variation in GS5 plays an important role in regulating grain size and yield in rice. *Nat Genet* **43**: 1266-1269
- Lijavetzky D, Carbonero P, Vicente-Carbajosa J (2003) Genome-wide comparative phylogenetic analysis of the rice and Arabidopsis Dof gene families. *Bmc Evolutionary Biology* **3**
- Liu C, Xu Z, Chua NH (1993) Auxin Polar Transport Is Essential for the Establishment of Bilateral Symmetry during Early Plant Embryogenesis. *Plant Cell* **5**: 621-630
- Liu LS, White MJ, MacRae TH (1999) Transcription factors and their genes in higher plants - Functional domains, evolution and regulation. *European Journal of Biochemistry* **262**: 247-257
- Liu Q, Zhang GY, Chen SY (2001) Structure and regulatory function of plant transcription factors. *Chinese Science Bulletin* **46**: 271-278
- Liu X, Zhang H, Zhao Y, Feng Z, Li Q, Yang HQ, Luan S, Li J, He ZH (2013) Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated ABI3 activation in Arabidopsis. *Proc Natl Acad Sci U S A* **110**: 15485-15490
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25**: 402-408
- Loftfield RB, Grover JW, Stephenson ML (1953) Possible Role of Proteolytic Enzymes in Protein Synthesis. *Nature* **171**: 1024-1025

Long TA, Benfey PN (2006) Transcription factors and hormones: new insights into plant cell differentiation. *Current Opinion in Cell Biology* **18**: 710-714

Lu CA, Ho TH, Ho SL, Yu SM (2002) Three novel MYB proteins with one DNA binding repeat mediate sugar and hormone regulation of alpha-amylase gene expression. *Plant Cell* **14**: 1963-1980

Lu M, Ying S, Zhang DF, Shi YS, Song YC, Wang TY, Li Y (2012) A maize stress-responsive NAC transcription factor, ZmSNAC1, confers enhanced tolerance to dehydration in transgenic Arabidopsis. *Plant Cell Rep* **31**: 1701-1711

Lu XD, Chen DJ, Shu DF, Zhang Z, Wang WX, Klukas C, Chen LL, Fan YL, Chen M, Zhang CY (2013) The Differential Transcription Network between Embryo and Endosperm in the Early Developing Maize Seed. *Plant Physiology* **162**: 440-455

Luang S, Cho JI, Mahong B, Opassiri R, Akiyama T, Phasai K, Komvongsa J, Sasaki N, Hua YL, Matsuba Y, Ozeki Y, Jeon JS, Cairns JRK (2013) Rice Os9BGlu31 Is a Transglucosidase with the Capacity to Equilibrate Phenylpropanoid, Flavonoid, and Phytohormone Glycoconjugates. *Journal of Biological Chemistry* **288**: 10111-10123

Ludwig-Muller J (2011) Auxin conjugates: their role for plant development and in the evolution of land plants. *J Exp Bot* **62**: 1757-1773

Luo M, Dennis ES, Berger F, Peacock WJ, Chaudhury A (2005) MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 17531-17536

-M-

Marshall E, Costa LM, Gutierrez-Marcos J (2011) Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. *J Exp Bot* **62**: 1677-1686

Marzabal P, Gas E, Fontanet P, Vicente-Carbajosa J, Torrent M, Ludevid MD (2008) The maize Dof protein PBF activates transcription of gamma-zein during maize seed development. *Plant Molecular Biology* **67**: 441-454

Mashiguchi K, Tanaka K, Sakai T, Sugawara S, Kawaide H, Natsume M, Hanada A, Yaeno T, Shirasu K, Yao H, McSteen P, Zhao Y, Hayashi K, Kamiya Y, Kasahara H (2011) The main auxin biosynthesis pathway in Arabidopsis. *Proc Natl Acad Sci U S A* **108**: 18512-18517

Mazzucotelli E, Mastrangelo AA, Crosatti C, Guerra D, Stanca AM, Cattivelli L (2008) Abiotic stress response in plants: When post-transcriptional and post-translational regulations control transcription. *Plant Science* **174**: 420-431

McCallum CM, Comai L, Greene EA, Henikoff S (2000) Targeting induced local lesions in genomes (TILLING) for plant functional genomics. *Plant Physiology* **123**: 439-442

- Melkus G, Rolletschek H, Radchuk R, Fuchs J, Rutten T, Wobus U, Altmann T, Jakob P, Borisjuk L** (2009) The metabolic role of the legume endosperm: a noninvasive imaging study. *Plant Physiol* **151**: 1139-1154
- Mena M, Vicente-Carbajosa J, Schmidt RJ, Carbonero P** (1998) An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamins-box of a native B-hordein promoter in barley endosperm. *Plant Journal* **16**: 53-62
- Mishra BS, Singh M, Aggrawal P, Laxmi A** (2009) Glucose and auxin signaling interaction in controlling *Arabidopsis thaliana* seedlings root growth and development. *PLoS One* **4**: e4502
- Mittasch J, Strack D, Milkowski C** (2007) Secondary product glycosyltransferases in seeds of *Brassica napus*. *Planta* **225**: 515-522
- Miyawaki K, Matsumoto-Kitano M, Kakimoto T** (2004) Expression of cytokinin biosynthetic isopentenyltransferase genes in *Arabidopsis*: tissue specificity and regulation by auxin, cytokinin, and nitrate. *Plant Journal* **37**: 128-138
- Mizutani M, Ohta D** (2010) Diversification of P450 Genes During Land Plant Evolution. *Annual Review of Plant Biology*, Vol 61 **61**: 291-315
- Moller B, Weijers D** (2009) Auxin control of embryo patterning. *Cold Spring Harb Perspect Biol* **1**: a001545
- Mondal P, Dey N, Dash AK, Chatterjee A, Satin BB, Panda B, Maiti IB, Sabat SC** (2007) Structural and functional analysis of rice catalase-B gene promoter: Presence of Dof and CAAT binding site. *Plant Molecular Biology Reporter* **25**: 71-82
- Moreau D, Salon C, Munier-Jolain N** (2006) Using a standard framework for the phenotypic analysis of *Medicago truncatula*: an effective method for characterizing the plant material used for functional genomics approaches. *Plant Cell Environ* **29**: 1087-1098
- Moreno-Risueno MA, Diaz I, Carrillo L, Fuentes R, Carbonero P** (2007) The HvDOF19 transcription factor mediates the abscisic acid-dependent repression of hydrolase genes in germinating barley aleurone. *Plant Journal* **51**: 352-365
- Moreno-Risueno MA, Martinez M, Vicente-Carbajosa J, Carbonero P** (2007) The family of DOF transcription factors: from green unicellular algae to vascular plants. *Molecular Genetics and Genomics* **277**: 379-390
- Murmu J, Bush MJ, DeLong C, Li S, Xu M, Khan M, Malcolmson C, Fobert PR, Zachgo S, Hepworth SR** (2010) *Arabidopsis* basic leucine-zipper transcription factors TGA9 and TGA10 interact with floral glutaredoxins ROXY1 and ROXY2 and are redundantly required for anther development. *Plant Physiol* **154**: 1492-1504

-N-

Nagasawa N, Hibara K, Heppard EP, Vander Velden KA, Luck S, Beatty M, Nagato Y, Sakai H (2013) GIANT EMBRYO encodes CYP78A13, required for proper size balance between embryo and endosperm in rice. *Plant J* **75**: 592-605

Nakano T, Suzuki K, Ohtsuki N, Tsujimoto Y, Fujimura T, Shinshi H (2006) Identification of genes of the plant-specific transcription-factor families cooperatively regulated by ethylene and jasmonate in *Arabidopsis thaliana*. *Journal of Plant Research* **119**: 407-413

Nardmann J, Zimmermann R, Durantini D, Kranz E, Werr W (2007) WOX gene phylogeny in Poaceae: a comparative approach addressing leaf and embryo development. *Mol Biol Evol* **24**: 2474-2484

Nayar S, Sharma R, Tyagi AK, Kapoor S (2013) Functional delineation of rice MADS29 reveals its role in embryo and endosperm development by affecting hormone homeostasis. *Journal of Experimental Botany* **64**: 4239-4253

Neff MM, Neff JD, Chory J, Pepper AE (1998) dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. *Plant Journal* **14**: 387-392

Negi J, Moriwaki K, Konishi M, Yokoyama R, Nakano T, Kusumi K, Hashimoto-Sugimoto M, Schroeder Julian I, Nishitani K, Yanagisawa S, Iba K (2013) A Dof Transcription Factor, SCAP1, Is Essential for the Development of Functional Stomata in *Arabidopsis*. *Current biology : CB* **23**: 479-484

Nemecek T, von Richthofen J-S, Dubois G, Casta P, Charles R, Pahl H (2008) Environmental impacts of introducing grain legumes into European crop rotations. *European Journal of Agronomy* **28**: 380-393

Nesi N, Debeaujon I, Jond C, Pelletier G, Caboche M, Lepiniec L (2000) The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and BAN genes in *Arabidopsis* siliques. *Plant Cell* **12**: 1863-1878

Nicolas P, Lecourieux D, Gomès E, Delrot S, Lecourieux F (2013) The grape berry-specific basic helix-loop-helix transcription factor VvCEB1 affects cell size. *Journal of Experimental Botany*

Noguero M, Atif RM, Ochatt S, Thompson RD (2013) The role of the DNA-binding One Zinc Finger (DOF) transcription factor family in plants. *Plant Science* **209**: 32-45

-P-

Paek NC, Lee BM, Bai DG, Smith JD (1998) Inhibition of germination gene expression by Viviparous-1 and ABA during maize kernel development. *Molecules and Cells* **8**: 336-342

- Palma JM, Sandalio LM, Javier Corpas F, Romero-Puertas MaC, McCarthy I, del R  o LA** (2002) Plant proteases, protein degradation, and oxidative stress: role of peroxisomes. *Plant Physiology and Biochemistry* **40**: 521-530
- Pang Y, Peel GJ, Sharma SB, Tang Y, Dixon RA** (2008) A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of *Medicago truncatula*. *Proc Natl Acad Sci U S A* **105**: 14210-14215
- Paolis Ad, Sabatini S, Pascalis Ld, Costantino P, Capone I** (1996) A rolB regulatory factor belongs to a new class of single zinc finger plant proteins. *Plant Journal* **10**: 215-223
- Papi M, Sabatini S, Altamura MM, Hennig L, Schafer E, Costantino P, Vittorioso P** (2002) Inactivation of the phloem-specific dof zinc finger gene DAG1 affects response to light and integrity of the testa of *Arabidopsis* seeds. *Plant Physiology* **128**: 411-417
- Papi M, Sabatini S, Bouchez D, Camilleri C, Costantino P, Vittorioso P** (2000) Identification and disruption of an *Arabidopsis* zinc finger gene controlling seed germination. *Genes & Development* **14**: 28-33
- Park DH, Lim PO, Kim JS, Cho DS, Hong SH, Nam HG** (2003) The *Arabidopsis* COG1 gene encodes a Dof domain transcription factor and negatively regulates phytochrome signaling. *Plant Journal* **34**: 161-171
- Pattison RJ, Catala C** (2012) Evaluating auxin distribution in tomato (*Solanum lycopersicum*) through an analysis of the PIN and AUX/LAX gene families. *Plant Journal* **70**: 585-598
- Pelzer E, Bazot M, Makowski D, Corre-Hellou G, Naudin C, Al Rifai M, Baranger E, Bedoussac L, Biarnes V, Boucheny P, Carrouee B, Dorvillez D, Foissy D, Gaillard B, Guichard L, Mansard M-C, Omon B, Prieur L, Yvergnaux M, Justes E, Jeuffroy M-H** (2012) Pea-wheat intercrops in low-input conditions combine high economic performances and low environmental impacts. *European Journal of Agronomy* **40**: 39-53
- Penfield S, Josse EM, Kannangara R, Gilday AD, Halliday KJ, Graham IA** (2005) Cold and light control seed germination through the bHLH transcription factor SPATULA. *Current Biology* **15**: 1998-2006
- Pignocchi C, Minns GE, Nesi N, Koumproglou R, Kitsios G, Benning C, Lloyd CW, Doonan JH, Hills MJ** (2009) ENDOSPERM DEFECTIVE1 Is a Novel Microtubule-Associated Protein Essential for Seed Development in *Arabidopsis*. *Plant Cell* **21**: 90-105
- Plesch G, Ehrhardt T, Mueller-Roeber B** (2001) Involvement of TAAAG elements suggests a role for Dof transcription factors in guard cell-specific gene expression. *Plant Journal* **28**: 455-464
- Ploense SE, Wu MF, Nagpal P, Reed JW** (2009) A gain-of-function mutation in IAA18 alters *Arabidopsis* embryonic apical patterning. *Development* **136**: 1509-1517
- Purugganan MD, Wessler SR** (1994) Molecular evolution of the plant R regulatory gene family. *Genetics* **138**: 849-854

-Q-

Quan R, Hu S, Zhang Z, Zhang H, Zhang Z, Huang R (2010) Overexpression of an ERF transcription factor TSRF1 improves rice drought tolerance. *Plant Biotechnol J* **8**: 476-488

-R-

Rajjou L, Duval M, Gallardo K, Catusse J, Bally J, Job C, Job D (2012) Seed germination and vigor. *Annu Rev Plant Biol* **63**: 507-533

Ramsay NA, Glover BJ (2005) MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. *Trends in Plant Science* **10**: 63-70

Reiser L, Fischer RL (1993) The Ovule and the Embryo Sac. *Plant Cell* **5**: 1291-1301

Repetto O, Rogniaux H, Firnhaber C, Zuber H, Kuster H, Larre C, Thompson R, Gallardo K (2008) Exploring the nuclear proteome of *Medicago truncatula* at the switch towards seed filling. *Plant Journal* **56**: 398-410

Richardt S, Timmerhaus G, Lang D, Qudeimat E, Correa LGG, Reski R, Rensing SA, Frank W (2010) Microarray analysis of the moss *Physcomitrella patens* reveals evolutionarily conserved transcriptional regulation of salt stress and abscisic acid signalling. *Plant Molecular Biology* **72**: 27-45

Rizza A, Boccaccini A, Lopez-Vidriero I, Costantino P, Vittorioso P (2011) Inactivation of the ELIP1 and ELIP2 genes affects *Arabidopsis* seed germination. *New Phytologist* **190**: 896-905

Rogers LA, Dubos C, Surman C, Willment J, Cullis IF, Mansfield SD, Campbell MM (2005) Comparison of lignin deposition in three ectopic lignification mutants. *New Phytologist* **168**: 123-140

Rose RJ (2008) *Medicago truncatula* as a model for understanding plant interactions with other organisms, plant development and stress biology: past, present and future. *Functional Plant Biology* **35**: 253-264

Rueda-Lopez M, Crespillo R, Canovas FM, Avila C (2008) Differential regulation of two glutamine synthetase genes by a single Dof transcription factor. *Plant Journal* **56**: 73-85

Rueda-Romero P, Barrero-Sicilia C, Gomez-Cadenas A, Carbonero P, Onate-Sanchez L (2012) *Arabidopsis thaliana* DOF6 negatively affects germination in non-after-ripened seeds and interacts with TCP14. *Journal of Experimental Botany* **63**: 1937-1949

-S-

Santos LA, de Souza SR, Fernandes MS (2012) OsDof25 expression alters carbon and nitrogen metabolism in *Arabidopsis* under high N-supply. *Plant Biotechnology Reports* **6**: 327-337

- Scherbak N, Ala-Haivala A, Brosche M, Bower N, Strid H, Gittins JR, Grahn E, Eriksson LA, Strid A (2011) The pea SAD short-chain dehydrogenase/reductase: quinone reduction, tissue distribution, and heterologous expression. *Plant Physiol* **155**: 1839-1850
- Schnabel EL, Frugoli J (2004) The PIN and LAX families of auxin transport genes in *Medicago truncatula*. *Mol Genet Genomics* **272**: 420-432
- Schneidereit A, Imlau A, Sauer N (2008) Conserved cis-regulatory elements for DNA-binding-with-one-finger and homeo-domain-leucine-zipper transcription factors regulate companion cell-specific expression of the *Arabidopsis thaliana* SUCROSE TRANSPORTER 2 gene. *Planta* **228**: 651-662
- Schruff MC, Spielman M, Tiwari S, Adams S, Fenby N, Scott RJ (2006) The AUXIN RESPONSE FACTOR 2 gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. *Development* **133**: 251-261
- Schwechheimer C, Bevan M (1998) The regulation of transcription factor activity in plants. *Trends in Plant Science* **3**: 378-383
- Scott RJ, Spielman M, Bailey J, Dickinson HG (1998) Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* **125**: 3329-3341
- Serrani JC, Carrera E, Ruiz-Rivero O, Gallego-Giraldo L, Peres LE, Garcia-Martinez JL (2010) Inhibition of auxin transport from the ovary or from the apical shoot induces parthenocarpic fruit-set in tomato mediated by gibberellins. *Plant Physiol* **153**: 851-862
- Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, Muehlbauer GJ, Nelson RT, Grant D, Specht JE, Graham MA, Cannon SB, May GD, Vance CP, Shoemaker RC (2010) RNA-Seq Atlas of Glycine max: a guide to the soybean transcriptome. *BMC Plant Biol* **10**: 160
- Shaw LM, McIntyre CL, Gresshoff PM, Xue G-P (2009) Members of the Dof transcription factor family in *Triticum aestivum* are associated with light-mediated gene regulation. *Functional & Integrative Genomics* **9**: 485-498
- Shigyo M, Tabei N, Yoneyama T, Yanagisawa S (2007) Evolutionary processes during the formation of the plant-specific Dof transcription factor family. *Plant and Cell Physiology* **48**: 179-185
- Shimofurutani N, Kisu Y, Suzuki M, Esaka M (1998) Functional analyses of the Dof domain, a zinc finger DNA-binding domain, in a pumpkin DNA-binding protein AOBP. *FEBS Lett* **430**: 251-256
- Siddique KHM, Johansen C, Turner NC, Jeuffroy M-H, Hashem A, Sakar D, Gan Y, Alghamdi SS (2012) Innovations in agronomy for food legumes. A review. *Agronomy for Sustainable Development* **32**: 45-64
- Sigrist CJA, Cerutti L, Hulo N, Gattiker A, Falquet L, Pagni M, Bairoch A, Bucher P (2002) PROSITE: a documented database using patterns and profiles as motif descriptors. *Briefings in bioinformatics* **3**: 265-274

Sigrist CJA, de Castro E, Cerutti L, Cucho BA, Hulo N, Bridge A, Bougueleret L, Xenarios I (2012) New and continuing developments at PROSITE. *Nucleic acids research* **41**: D344-347

Skirycz A, Jozefczuk S, Stobiecki M, Muth D, Zanon MI, Witt I, Mueller-Roeber B (2007) Transcription factor AtDOF4;2 affects phenylpropanoid metabolism in *Arabidopsis thaliana*. *New Phytol* **175**: 425-438

Skirycz A, Radziejowski A, Busch W, Hannah MA, Czeszejko J, Kwasniewski M, Zanon MI, Lohmann JU, De Veylder L, Witt I, Mueller-Roeber B (2008) The DOF transcription factor OBP1 is involved in cell cycle regulation in *Arabidopsis thaliana*. *Plant J* **56**: 779-792

Skirycz A, Reichelt M, Burow M, Birkemeyer C, Rolcik J, Kopka J, Zanon MI, Gershenzon J, Strnad M, Szopa J, Mueller-Roeber B, Witt I (2006) DOF transcription factor AtDof1.1 (OBP2) is part of a regulatory network controlling glucosinolate biosynthesis in *Arabidopsis*. *Plant Journal* **47**: 10-24

Smyth GK, Michaud JL, Scott HS (2005) Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* **21**: 2067-2075

Song YH, Smith RW, To BJ, Millar AJ, Imaizumi T (2012) FKF1 Conveys Timing Information for CONSTANS Stabilization in Photoperiodic Flowering. *Science* **336**: 1045-1049

Stepanova AN, Robertson-Hoyt J, Yun J, Benavente LM, Xie D-Y, Dolezal K, Schlereth A, Juergens G, Alonso JM (2008) TAA1-mediated auxin biosynthesis is essential for hormone crosstalk and plant development. *Cell* **133**: 177-191

Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, Goldberg RB, Harada JJ (2001) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. *Proc Natl Acad Sci U S A* **98**: 11806-11811

Sugiyama T, Ishida T, Tabei N, Shigyo M, Konishi M, Yoneyama T, Yanagisawa S (2012) Involvement of PpDof1 transcriptional repressor in the nutrient condition-dependent growth control of protonemal filaments in *Physcomitrella patens*. *Journal of Experimental Botany* **63**: 3185-3197

Sun XD, Shantharaj D, Kang XJ, Ni M (2010) Transcriptional and hormonal signaling control of *Arabidopsis* seed development. *Current Opinion in Plant Biology* **13**: 611-620

Sundberg E, Ostergaard L (2009) Distinct and dynamic auxin activities during reproductive development. *Cold Spring Harb Perspect Biol* **1**: a001628

-T-

Tadege M, Wen JQ, He J, Tu HD, Kwak Y, Eschstruth A, Cayrel A, Endre G, Zhao PX, Chabaud M, Ratet P, Mysore KS (2008) Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. *Plant Journal* **54**: 335-347

- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725-2729
- Tanaka H, Onouchi H, Kondo M, Hara-Nishimura I, Nishimura M, Machida C, Machida Y (2001) A subtilisin-like serine protease is required for epidermal surface formation in *Arabidopsis* embryos and juvenile plants. *Development* **128**: 4681-4689
- Tanaka M, Takahata Y, Nakayama H, Nakatani M, Tahara M (2009) Altered carbohydrate metabolism in the storage roots of sweetpotato plants overexpressing the SRF1 gene, which encodes a Dof zinc finger transcription factor. *Planta* **230**: 737-746
- Tang W, Wang W, Chen D, Ji Q, Jing Y, Wang H, Lin R (2013) Transposase-derived proteins FHY3/FAR1 interact with PHYTOCHROME-INTERACTING FACTOR1 to regulate chlorophyll biosynthesis by modulating HEMB1 during deetiolation in *Arabidopsis*. *Plant Cell* **24**: 1984-2000
- Tesfaye M, Silverstein KA, Nallu S, Wang L, Botanga CJ, Gomez SK, Costa LM, Harrison MJ, Samac DA, Glazebrook J, Katagiri F, Gutierrez-Marcos JF, Vandenbosch KA (2013) Spatio-temporal expression patterns of *Arabidopsis thaliana* and *Medicago truncatula* defensin-like genes. *PLoS One* **8**: e58992
- Thangstad OP, Gilde B, Chadchawan S, Seem M, Husebye H, Bradley D, Bones AM (2004) Cell specific, cross-species expression of myrosinases in *Brassica napus*, *Arabidopsis thaliana* and *Nicotiana tabacum*. *Plant Molecular Biology* **54**: 597-611
- Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant Journal* **37**: 914-939
- Tivendale ND, Davies NW, Molesworth PP, Davidson SE, Smith JA, Lowe EK, Reid JB, Ross JJ (2010) Reassessing the role of N-hydroxytryptamine in auxin biosynthesis. *Plant Physiol* **154**: 1957-1965
- Tremousaygue D, Garnier L, Bardet C, Dabos P, Herve C, Lescure B (2003) Internal telomeric repeats and 'TCP domain' protein-binding sites co-operate to regulate gene expression in *Arabidopsis thaliana* cycling cells. *Plant J* **33**: 957-966
- Tsujimoto-Inui Y, Naito Y, Sakurai N, Suzuki H, Sasaki R, Takahashi H, Ohtsuki N, Nakano T, Yanagisawa S, Shibata D, Uchimiya H, Shinshi H, Suzuki K (2009) Functional genomics of the Dof transcription factor family genes in suspension-cultured cells of *Arabidopsis thaliana*. *Plant Biotechnology* **26**: 15-28

-U-

- Umemura Y, Ishiduka T, Yamamoto R, Esaka M (2004) The Dof domain, a zinc finger DNA-binding domain conserved only in higher plants, truly functions as a Cys2/Cys2 Zn finger domain. *Plant J* **37**: 741-749

Usadel B, Nagel A, Steinhauser D, Gibon Y, Blasing OE, Redestig H, Sreenivasulu N, Krall L, Hannah MA, Poree F, Fernie AR, Stitt M (2006) PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. *BMC Bioinformatics* **7**: 535

-V-

Vandecasteele C, Teulat-Merah B, Morere-Le Paven M-C, Leprince O, Vu BL, Viau L, Ledroit L, Pelletier S, Payet N, Satour P, Lebras C, Gallardo K, Huguet T, Limami AM, Prosperi J-M, Buitink J (2011) Quantitative trait loci analysis reveals a correlation between the ratio of sucrose/raffinose family oligosaccharides and seed vigour in *Medicago truncatula*. *Plant Cell and Environment* **34**: 1473-1487

Vanneste S, Friml J (2009) Auxin: a trigger for change in plant development. *Cell* **136**: 1005-1016

Venglat P, Xiang D, Qiu S, Stone SL, Tibiche C, Cram D, Alting-Mees M, Nowak J, Cloutier S, Deyholos M, Bekkaoui F, Sharpe A, Wang E, Rowland G, Selvaraj G, Datla R (2011) Gene expression analysis of flax seed development. *BMC Plant Biol* **11**: 74

Venglat P, Xiang D, Yang H, Wan L, Tibiche C, Ross A, Wang E, Selvaraj G, Datla R (2013) Gene expression profiles during embryo development in *Brassica napus*. *Plant Breeding* **132**: 514-522

Verdier J, Kakar K, Gallardo K, Le Signor C, Aubert G, Schlereth A, Town CD, Udvardi MK, Thompson RD (2008) Gene expression profiling of *M-truncatula* transcription factors identifies putative regulators of grain legume seed filling. *Plant Molecular Biology* **67**: 567-580

Verdier J, Zhao J, Torres-Jerez I, Ge S, Liu C, He X, Mysore KS, Dixon RA, Udvardi MK (2012) MtPAR MYB transcription factor acts as an on switch for proanthocyanidin biosynthesis in *Medicago truncatula*. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 1766-1771

Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ (1997) A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 7685-7690

Vierstra RD (1996) Proteolysis in plants: Mechanisms and functions. *Plant Molecular Biology* **32**: 275-302

Vogt T, Jones P (2000) Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends in Plant Science* **5**: 380-386

Voisin A-S, Guéguen J, Huyghe C, Jeuffroy M-Hln, Magrini M-B, Meynard J-M, Mougél C, Pellerin S, Pelzer E (2013) Legumes for feed, food, biomaterials and bioenergy in Europe: a review. *Agronomy for Sustainable Development*: 1-20

Vriezen WH, Feron R, Maretto F, Keijman J, Mariani C (2008) Changes in tomato ovary transcriptome demonstrate complex hormonal regulation of fruit set. *New Phytologist* **177**: 60-76

-W-

Wang GF, Wang H, Zhu J, Zhang J, Zhang XW, Wang F, Tang YP, Mei B, Xu ZK, Song RT (2010) An expression analysis of 57 transcription factors derived from ESTs of developing seeds in Maize (*Zea mays*). *Plant Cell Reports* **29**: 545-559

Wang HL, Grusak MA (2005) Structure and development of *Medicago truncatula* pod wall and seed coat. *Ann Bot* **95**: 737-747

Wang HW, Zhang B, Hao YJ, Huang J, Tian AG, Liao Y, Zhang JS, Chen SY (2007) The soybean Dof-type transcription factor genes, GmDof4 and GmDof11, enhance lipid content in the seeds of transgenic *Arabidopsis* plants. *Plant J* **52**: 716-729

Wang M, Liu C, Li S, Zhu D, Zhao Q, Yu J (2013) Improved Nutritive Quality and Salt Resistance in Transgenic Maize by Simultaneously Overexpression of a Natural Lysine-Rich Protein Gene, SBgLR, and an ERF Transcription Factor Gene, TSRF1. *Int J Mol Sci* **14**: 9459-9474

Wang X-D, Song Y, Sheahan MB, Garg ML, Rose RJ (2011) From embryo sac to oil and protein bodies: embryo development in the model legume *Medicago truncatula*. *New Phytologist* **193**: 327-338

Ward JM, Cufr CA, Denzel MA, Neff MM (2005) The dof transcription factor OBP3 modulates phytochrome and cryptochrome signaling in *Arabidopsis*. *Plant Cell* **17**: 475-485

Washio K (2001) Identification of Dof proteins with implication in the gibberellin-regulated expression of a peptidase gene following the germination of rice grains. *Biochimica Et Biophysica Acta-Gene Structure and Expression* **1520**: 54-62

Washio K (2003) Functional dissections between GAMYB and dof transcription factors suggest a role for protein-protein associations in the gibberellin-mediated expression of the RAMy1A gene in the rice aleurone. *Plant Physiology* **133**: 850-863

Weber H, Borisjuk L, Wobus U (1997) Sugar import and metabolism during seed development. *Trends in Plant Science* **2**: 169-174

Weber H, Borisjuk L, Wobus U (2005) Molecular physiology of legume seed development. *Annual Review of Plant Biology* **56**: 253-279

Wei P-C, Tan F, Gao X-Q, Zhang X-Q, Wang G-Q, Xu H, Li L-J, Chen J, Wang X-C (2010) Overexpression of AtDOF4.7, an *Arabidopsis* DOF Family Transcription Factor, Induces Floral Organ Abcission Deficiency in *Arabidopsis*. *Plant Physiology* **153**: 1031-1045

Won C, Shen X, Mashiguchi K, Zheng Z, Dai X, Cheng Y, Kasahara H, Kamiya Y, Chory J, Zhao Y (2011) Conversion of tryptophan to indole-3-acetic acid by TRYPTOPHAN AMINOTRANSFERASES OF ARABIDOPSIS and YUCCAs in *Arabidopsis*. *Proc Natl Acad Sci U S A* **108**: 18518-18523

Woodward AW, Bartel B (2005) Auxin: regulation, action, and interaction. *Ann Bot* **95**: 707-735

Wu X, Chory J, Weigel D (2007) Combinations of WOX activities regulate tissue proliferation during Arabidopsis embryonic development. *Dev Biol* **309**: 306-316

-X-

Xiang D, Venglat P, Tibiche C, Yang H, Risseuw E, Cao Y, Babic V, Cloutier M, Keller W, Wang E, Selvaraj G, Datla R (2011) Genome-wide analysis reveals gene expression and metabolic network dynamics during embryo development in Arabidopsis. *Plant Physiol* **156**: 346-356

-Y-

Yamaguchi-Shinozaki K, Shinozaki K (2006) Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *In Annual Review of Plant Biology*, Vol 57, pp 781-803

Yamamoto MP, Onodera Y, Touno SM, Takaiwa F (2006) Synergism between RPB1 Dof and RISBZ1 bZIP activators in the regulation of rice seed expression genes. *Plant Physiol* **141**: 1694-1707

Yanagisawa S (1997) Dof DNA-binding domains of plant transcription factors contribute to multiple protein-protein interactions. *Eur J Biochem* **250**: 403-410

Yanagisawa S (2000) Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize. *Plant Journal* **21**: 281-288

Yanagisawa S (2001) The transcriptional activation domain of the plant-specific Dof1 factor functions in plant, animal, and yeast cells. *Plant and Cell Physiology* **42**: 813-822

Yanagisawa S (2002) The Dof family of plant transcription factors. *Trends in Plant Science* **7**: 555-560

Yanagisawa S, Akiyama A, Kisaka H, Uchimiya H, Miwa T (2004) Metabolic engineering with Dof1 transcription factor in plants: Improved nitrogen assimilation and growth under low-nitrogen conditions. *Proceedings of the National Academy of Sciences of the United States of America* **101**: 7833-7838

Yanagisawa S, Izui K (1993) Molecular-Cloning of 2 DNA-Binding Proteins of Maize That Are Structurally Different but Interact with the Same Sequence Motif. *Journal of Biological Chemistry* **268**: 16028-16036

Yanagisawa S, Schmidt RJ (1999) Diversity and similarity among recognition sequences of Dof transcription factors. *Plant Journal* **17**: 209-214

Yanagisawa S, Sheen J (1998) Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression. *Plant Cell* **10**: 75-89

Yang J, Yang MF, Zhang WP, Chen F, Shen SH (2011) A putative flowering-time-related Dof transcription factor gene, JcDof3, is controlled by the circadian clock in *Jatropha curcas*. *Plant Science* **181**: 667-674

Yang S, Johnston N, Talideh E, Mitchell S, Jeffree C, Goodrich J, Ingram G (2008) The endosperm-specific ZHOUP1 gene of *Arabidopsis thaliana* regulates endosperm breakdown and embryonic epidermal development. *Development* **135**: 3501-3509

Yang WB, Gao MJ, Yin X, Liu JY, Xu YH, Zeng LJ, Li Q, Zhang SB, Wang JM, Zhang XM, He ZH (2013) Control of Rice Embryo Development, Shoot Apical Meristem Maintenance, and Grain Yield by a Novel Cytochrome P450. *Molecular Plant* **6**: 1945-1960

Yang X, Tuskan GA, Cheng Z-M (2006) Divergence of the Dof gene families in poplar, *Arabidopsis*, and rice suggests multiple modes of gene evolution after duplication. *Plant Physiology* **142**: 820-830

Young ND, Debelle F, Oldroyd GED, Geurts R, Cannon SB, Udvardi MK, Benedito VA, Mayer KFX, Gouzy J, Schoof H, Van de Peer Y, Proost S, Cook DR, Meyers BC, Spannagl M, Cheung F, De Mita S, Krishnakumar V, Gundlach H, Zhou SG, Mudge J, Bharti AK, Murray JD, Naoumkina MA, Rosen B, Silverstein KAT, Tang HB, Rombauts S, Zhao PX, Zhou P, Barbe V, Bardou P, Bechner M, Bellec A, Berger A, Berges H, Bidwell S, Bisseling T, Choise N, Couloux A, Denny R, Deshpande S, Dai XB, Doyle JJ, Dudez AM, Farmer AD, Fouteau S, Franken C, Gibelin C, Gish J, Goldstein S, Gonzalez AJ, Green PJ, Hallab A, Hartog M, Hua A, Humphray SJ, Jeong DH, Jing Y, Jocker A, Kenton SM, Kim DJ, Klee K, Lai HS, Lang CT, Lin SP, Macmil SL, Magdelenat G, Matthews L, McCorrison J, Monaghan EL, Mun JH, Najjar FZ, Nicholson C, Noirot C, O'Brien M, Paule CR, Poulain J, Prion F, Qin BF, Qu CM, Retzel EF, Riddle C, Sallet E, Samain S, Samson N, Sanders I, Saurat O, Scarpelli C, Schiex T, Segurens B, Severin AJ, Sherrier DJ, Shi RH, Sims S, Singer SR, Sinharoy S, Sterck L, Viollet A, Wang BB, Wang KQ, Wang MY, Wang XH, Warfsmann J, Weissenbach J, White DD, White JD, Wiley GB, Wincker P, Xing YB, Yang LM, Yao ZY, Ying F, Zhai JX, Zhou LP, Zuber A, Denarie J, Dixon RA, May GD, Schwartz DC, Rogers J, Quetier F, Town CD, Roe BA (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**: 520-524

-Z-

Zhang B, Chen W, Foley RC, Buttner M, Singh KB (1995) Interactions between distinct types of DNA binding proteins enhance binding to ocs element promoter sequences. *Plant Cell* **7**: 2241-2252

Zhang H, Jin J, Tang L, Zhao Y, Gu X, Gao G, Luo J (2011) PlantTFDB 2.0: update and improvement of the comprehensive plant transcription factor database. *Nucleic Acids Res* **39**: D1114-1117

Zhang Y, Cao G, Qu L, Gu H (2009) Involvement of an R2R3-MYB transcription factor gene AtMYB118 in embryogenesis in *Arabidopsis*. *Plant Cell Reports* **28**: 337-346

Zhang Y, Liang W, Shi J, Xu J, Zhang D (2013) MYB56 Encoding a R2R3 MYB Transcription Factor Regulates Seed Size in *Arabidopsis thaliana*. *Journal of Integrative Plant Biology* **55**: 1166-1178

Zhao Y, Christensen SK, Fankhauser C, Cashman JR, Cohen JD, Weigel D, Chory J (2001) A role for flavin monooxygenase-like enzymes in auxin biosynthesis. *Science* **291**: 306-309

Zheng SX, Xiao S, Chye ML (2012) The gene encoding *Arabidopsis* acyl-CoA-binding protein 3 is pathogen inducible and subject to circadian regulation. *Journal of Experimental Botany* **63**: 2985-3000

Zhou Y, Huang W, Liu L, Chen T, Zhou F, Lin Y (2013) Identification and functional characterization of a rice NAC gene involved in the regulation of leaf senescence. *BMC Plant Biology* **13**: 132

Zou X, Neuman D, Shen QJ (2008) Interactions of two transcriptional repressors and two transcriptional activators in modulating gibberellin signaling in aleurone cells. *Plant Physiol* **148**: 176-186

Zuber H, Noguero M, Le Signor C, Thompson R, Gallardo K (2012) Metabolic Specialization of Maternal and Filial Tissues. *In* GK Agrawal, R Rakwal, eds, *Seed Development: OMICS Technologies toward Improvement of Seed Quality and Crop Yield*. Springer Netherlands, pp 407-432